JOURNAL *of* EDUCATIONAL RESEARCH

Volume XXXIII FEBRUARY, 1940 Number 6

RELATIVE DIFFICULTY OF TYPES OF EXAMPLES IN DIVISION WITH TWO-FIGURE DIVISORS

LEO J. BRUECKNER and HARVEY O. MELBYE

University of Minnesota

Editor's note: Much excellent research has been conducted in recent years on the teaching of arithmetic. The authors offer a study of the relative difficulty of the various elements in long division with two place divisors for pupils in grades 5B to 6A.

In recent years there have been numerous investigations of the teaching of long division. There are available, for example, the detailed analyses of the skills in long division by Upton[1], Grossnickle[2], and others. Grossnickle[3] has shown that under carefully supervised conditions this process can be taught at the fourth grade level. Grossnickle has shown that teaching pupils to estimate quotient figures by either the increase-by-one method, or by the apparent method, produces about equally good results.

Probably the most influential study in long division which has yet been undertaken is that of the Illinois Committee of Seven[4]. This committee conducted an experiment on a large scale to determine the optimum point at which to teach the entire process to completion. Originally the committee proposed that the process of long division should be taught to pupils of the minimum mental age of about 12 years 6 months. In reports of the investigations of the committee it has been carefully indicated that the committee made no recommendation as to the points at which

[1] Upton, C. B. "Making Long Division Automatic," in *The Teaching of Arithmetic*, Tenth Yearbook of the National Council of Teachers of Mathematics (New York: Teacher College Press, Columbia University, 1935), pp. 251-289.

[2] Grossnickle, Foster E. "Classification of the Estimations in Two Methods of Finding the Quotient in Long Division," *Elementary School Journal*, XXXII (April, 1932), pp. 595-604.

[3] Grossnickle, Foster E. "An Experiment with Two Methods of Estimating the Quotient," *Elementary School Journal*, XXXVII (May, 1937), pp. 668-677.

[4] An extensive bibliography of the publications of the Committee of Seven is given at the end of the following article by C. Washburne, "The Values, Limitation, and Applications of the Findings of the Committee of Seven," *Journal of Educational Research*, XXIX (May, 1936), pp. 694-707.

the easier elements of the process can be taught effectively. This matte was left open for further study by others interested in it. Decisions a to where to introduce the easy steps in the process were left to loca groups of teachers and curriculum workers, whose judgments unfortu nately were unsupported by any scientific evidence as to the relativ difficulty of the various elements.

The general recommendation of the Committee of Seven has been attacked for a number of reasons[5], among them the inadequacy of the test used to measure achievement in long division which consisted o only ten examples, four of them with one figure quotients, the failure o the test to cover adequately all major elements and difficulties of the process, the unreliable measures of mental ability used, the possibility that the instructional materials used were not constructed as efficiently as they might have been with the information now available, and the lack of control of conditions under which the teaching was done.

The criterion set up, namely 80 percent mastery—equivalent to 20 percent of error—by 75 percent of the pupils on a retention test following a six weeks' period in which there was no practice on the process, has also been severely criticized on the ground that it does not represent any usual plan of determining mastery. The committee has replied that any school is quite free to set up some other criterion more acceptable to the group concerned. The degree of accuracy set up by the committee as criterion of mastery amounts to at least 60 percent for a class as a whole (assuming that three-fourths had scores of 80 percent and the remainder scores of 0 percent). The higher the percent of accuracy achieved by those who had at least 80 percent of mastery the higher would be the class average. This result would also be true the nearer those who were below the standard of 80 percent approached it. It thus appears that the criterion set up by the committee when translated into a percent of accuracy for the group as a whole would be about 75 percent. This is the criterion that will be employed throughout the present study as a suggested index of mastery to be used as a basis of gradation. The two indices are not strictly comparable, yet they enable us to make useful comparisons between the results of the study of the Committee of Seven and the present investigation.

The point has also been made by critics that even if the process of long division *can* be taught to pupils of the minimum mental age of 12 years 6 months, we have no right to assume that it *should* be taught them.

[5] Brownell, W. A. "Readiness and the Arithmetic Curriculum," *Elementary School Journal*, XXXVIII (January, 1938), pp. 344-354.

These critics hold that long division should be taught when it meets some vital need of the individual. They claim that no one has yet determined by systematic research the point at which the actual need of this process arises in the normal activities of children. Some of them would go so far as to delay the teaching of long division until late in the school, as late as grades 7 and 8, on the ground that it does not function in life until even later on. It appears that we need at least two additional types of research in long division, namely, studies of the need of long division in the lives of children, and studies of methods and materials for teaching the process to pupils at various stages of development so that the process can be taught successfully at various points in the curriculum when the need for mastering it may arise. Ideally the process should be taught when there is a definite need for it and by means of materials and methods adapted to the level of development of the individual learner. No one knows the results that would be secured by such a program. On the other hand, the fact that long division *can* be taught at any given point in the curriculum is not a valid reason for teaching it at that time. It is possible that the process could be mastered much more easily and permanently if it were delayed for one, or two, or even more years. No one knows. It may be that instead of teaching the whole process to completion at one time it would be desirable to arrange the teaching program in such a way that the learning would be spread over a number of years, the easier elements coming earlier in the program and the more difficult elements at a later time. It is with this problem of gradation that this article is concerned.

We wish to report the results of a study, made at the University of Minnesota, of the relative difficulty of the various elements in long division with two-place divisors for pupils in grades 5B to 6A in a typical school system. The rudiments of the process are taught in this system in grade 4A and the process is completed in grade 5B or early in grade 5A. The method of teaching used in the basic textbook is the increase-by-one method for all divisors ending in 6, 7, 8, and 9. The teachers of the classes concerned are all rated as successful and in the three schools concerned there are principals who are rated as superior. The schools are all located in average residential areas. About 600 pupils were tested and complete data were secured for 474 pupils.

Two tests in long division were administered. Test A, administered in three parts, consisted of *32* examples in each of which there was only a one place quotient figure, with or without a remainder. The divisors

covered all decades and the second figures in them were well distribute from 0 to 9. The pupils were told to express the remainders in t quotients as remainders, not as fractions, so as to eliminate any difficul due to reduction of fractions or the use of decimals. The examples we so constructed that all major combinations of skills and difficulties naming quotients were tested. The test included 20 examples in whi the apparent quotient is the true quotient, and 12 others in which t apparent quotient is 1, 2, 3, or more too large, when the apparent meth of estimating the quotient is used, or in which the divisors were tee The examples in the test are defined in Table I. Test B, administer in two parts, consisted of *36* examples involving two- and three-figu quotients, zero difficulties in quotients, remainders, and difficulties naming the kinds of quotients that were tested in Test A. The types a defined in Table IV. Thus, in each test there were items, usually two each subtype, that ranged from very simple to very difficult, or comple There was no time limit on any of the tests. The tests were given l in November of 1938.

To get a measure of the relative difficulty of each type of example was assumed that the percent of error by pupils who have been tau the process would be the best index, much as has been done in spelli and other subjects. This procedure is obviously not as acceptable as t determination of learning difficulty under carefully controlled condition as was done, for example, for addition combinations by Knight a Behrens[6]. The data here reported represent difficulty only in so far a the measure of the ability of the pupils to solve the examples after th have been taught the process is a measure of their difficulty. The da have been analyzed by half grades and by mental ages. The results f the three schools are combined in the tables that follow. Data were al secured as to the length of time required to solve each of the exampl to get a further index of the difficulty. These data are not yet availabl Preliminary analysis of the results indicates that there are very wi variations in the amount of time pupils require to work the differe examples.

In Table I are given the percents of error for each grade from 5B t 6A for the various types of examples. From the data in row (1) it ca be seen that the average error for all types of examples combined rang from 46.0 in grade 5B to 28.7 in grade 6B. There was a slight increas in percent of error from grade 6B to grade 6A. The proposed limit o 25 percent error set up as a standard for gradation was not reached b

e end of grade 6A. It is interesting to speculate on what the results ould have been for grades 7 and 8.

As has been shown, the examples can be grouped as of two types. y comparing the data in rows (2) and (3) in Table I it can be seen at the examples in which the apparent quotient is the true quotient e much easier for the pupils in all grades than the type in which there ust be a correction when estimating the quotient, since the apparent otient is too large. The difference between the means for the two ajor types (rows 2 and 3) is greatest in grade 5B, namely 39.2 percent, d least in grade 6A, namely, 22.9 percent.

TABLE I

PERCENTS OF ERROR BY GRADE ON EXAMPLES
WITH ONE-FIGURE QUOTIENTS GROUPED ACCORDING TO TYPE OF DIFFICULTY

	5B	5A	6B	6A	No. of Examples
No. of Cases	125	103	143	103	
(1) All types combined	46.0	43.8	28.7	31.9	32
(2) Apparent quotient is true quotient	16.3	19.6	10.8	14.3	20
(a) Range on examples	3.8 to 34.9	.9 to 36.2	0 to 21.1	.9 to 29.6	20
(3) Apparent quotient is not true quotient	55.5	50.9	34.0	37.2	12
(a) Is 1 less than apparent	55.4	47.0	38.6	38.9	2
(b) Type: 728÷75(9)*	60.8	53.0	39.5	40.7	2
(c) Type: 302÷38(7-8)**	70.1	64.2	44.7	47.7	2
(d) Quotient is 2 or 3 less than apparent	51.9	53.0	34.5	35.6	2
(e) Teens (74÷14)	36.4	34.9	17.1	18.0	2
(f) Teens (155÷19)	60.8	53.4	34.5	42.1	2
(g) Range on examples	35.7 to 71.3	33.6 to 67.2	16.5 to 50.0	15.7 to 50.0	12

(2)* Means that the quotient figure is 9 although it appears to be 10.
(3-b)** Means that the quotient figure is 7 or 8 although it appears to be 10 or more.

The differences in these percents of error are very striking in all grades. They show that there is a marked difference in the difficulty of these two major types of examples. The ranges given in rows 2a and 3g show that there is a wide variation in the difficulty of the various examples included in each of these major groupings. In grade 5B these two ranges do not overlap at all; they overlap only slightly in grades 5A and 6B; and they overlap considerably in grade 6A. These facts suggest that there is a clean cut difference in the difficulty of the two types of examples. It is obvious that a mean degree of error of 25 percent which we shall assume is a reasonable criterion for determining gradation of steps in any process can probably be achieved for examples

[1] Knight, F. B., and Behrens, M. *Learning the 100 Addition Combinations and the 100 Subtraction Combinations*, New York: Longmans, Green and Company, 1928.

of the first general type in grade 5B. For the other type this standard is obviously not reached by the end of grade 6A. At this point the percent of error is still 37.2, approximately equivalent to what is apparently the lowest possible limit under the basic Committee of Seven standard. It can also be seen from Table I that the percent of error on examples in which the apparent quotient is the true quotient is less in grade 5B than in grade 5A, while the reverse is true for the examples in which the apparent quotient is not the true quotient. This situation suggests a possible interference factor operating in grade 5A.

In Table I the data for the second general type are further broken down according to six sub-types, for each of which the mean percents of error for two examples of each sub-type are given. Each sub-type is described in the table. The easiest sub-type in all grades is the type, ten divided by teens (row 3e). The most difficult sub-type in all grades is the type (302÷38) in which the apparent quotient is 10 or more and in which the true quotient is two, three, or four less than the apparent quotient. For all sub-types the percents of error are very large. They decrease somewhat from grade 5B to grade 6B and then increase slightly in grade 6A. It is obvious that on the basis of the findings it would not be earlier than the seventh or eighth grade that one should ordinarily expect a degree of error of less than 25 percent, assuming that the results here secured are typical of results in general.

These data show conclusively, it seems to us, that one should not speak of a process of long division as if it were one general ability of a general type or level of difficulty. It is clearly made up of a wide range of combinations of skills of marked differences in difficulty. It seems evident that the two major types of examples, here considered, should not be taught at the same grade level.

The Committee of Seven recommended that the long division process should be taught to completion when the pupil has a mental age of 12 years 6 months. The validity of this recommendation will now be considered.

In Table II the findings for each type of example with one-figure quotients various sub-types are presented for a considerable range of mental ages. Mental ages were derived from the Otis Self-Administering Test of Mental Ability. In this table the results for pupils of each mental age for all four grades are combined. The data are arranged in the same form as in Table I for each mental age group.

According to the data in row (1) there is a general decrease in percents of error for all types combined from level to level of mental age

TABLE II

PERCENTS OF ERROR BY MENTAL AGE GROUPS ON EXAMPLES WITH ONE-FIGURE QUOTIENTS (GRADES 5B-6A COMBINED)

		Mental Ages in Months						
		96-107	108-119	120-131	132-143	144-155	156-167	168 or more
(1)	All types combined	60.1	61.6	34.2	26.1	24.8	15.3	12.0
(2)	Apparent quotient is true quotient	51.3	51.7	20.2	13.8	14.2	7.8	6.1
(3)	Apparent quotient is not true quotient	75.0	78.1	57.4	46.6	42.6	28.0	21.6
a.	Is 1 less than apparent	75.0	79.6	53.4	41.2	42.4	31.1	19.6
b.	Type: 728÷75 (9)	75.0	79.6	59.3	51.2	50.4	29.1	30.4
c.	Type: 302÷38 (7-8)	75.0	85.2	69.6	61.5	59.3	35.1	30.4
d.	2 or 3 less than apparent	75.0	74.1	57.7	45.4	38.5	25.0	20.6
e.	Teens (74÷14)	75.0	63.9	44.6	24.6	20.4	13.5	5.4
f.	Teens (155÷19)	75.0	87.0	59.8	55.7	44.2	33.1	22.8
	Number of cases	4	24	92	120	121	73	40

In the mental age group, 144-155 months, or from 12 years to 13 years, the mental age recommended by the Committee of Seven for teaching long division, we find that the degree of error for all types of example with one-figure quotients combined, is approximately 25 percent. However, when the total data are divided according to the two major types of quotient difficulties, we find very large differences in the percents of error for all mental age groups. Apparently the standard of 25 percent error is reached before the mental age of 120 months for the examples of the first major type, but it is not reached for the second major type until at the very earliest the mental age of 168 months, about four years later. Here again we have evidence that long division should not be regarded as a single general process to be taught as a major unit, but that it contains a number of different elements that differ very greatly in their difficulty for pupils. If one accepts the idea of "pacing" instruction according to the rate of child development, present practices of teaching all types of long division examples with two-figure divisors in one grade are clearly unsound and indefensible.

These points are even more clearly brought out by the data for the five different sub-types under the second major type. As was true when the data were arranged by grades, the type (tens-divided-by-teens) was the easiest at all mental age levels and the type (302÷38) was the most difficult at all mental age levels. With the exception of the tens-divided-by-teens type all of the sub-types were so difficult that a mental age of at least 156-167 months was required to reduce the percent of error to approximately 25 percent. The range in percents of error for the various sub-types is small for each of the mental age groups. The percents in

general tend to decrease from left to right in the table, showing a positive relationship between mental maturity and success in mastering each sub-type. There appears to be little difference in the difficulty of the two sub-types in which the apparent quotient figure either must be reduced (1) or must be reduced (2), three (3), or more. There is a marked difference in the difficulty of the two sub-types involving teens, often called "the long division demons." The sub-type, tens-divided-by-teens, is clearly much easier than the sub-type, hundreds-divided-by-teens. The latter type certainly is a "demon."

In passing it should be pointed out that of the ten examples included in the Committee of Seven retention test four had one-figure quotients, and in three of the four the apparent quotient was the true quotient. Thus in only one of the four examples was a correction necessary. These facts raise a serious question as to the validity of the test as a measure of ability in long division since it was heavily loaded with easy items.

To bring out the way in which the results are affected when mental age level and grade level are considered together the results for three examples are shown in Table III. The data are arranged by grade for pupils of given mental ages in the grade. For example, there were pupils of the mental age 132-143 months in all grades. On the first example given, the percents of errors for pupils of that mental age in the four grades included were 5.3, 5.4, 10.3, and 6.3 respectively. While on the whole the pupils in grade 6A were about two years older chronologically than those of the same mental age in grade 5B, yet the percents of error for all groups of this mental age considered by grade level were strikingly alike. An examination of the results for the other examples and for the other mental age levels will show that the general tendency throughout is for these percents to remain quite constant from grade to grade for any given mental age for all examples. There are several exceptions to this general rule, for some of the other examples, due probably to the limited number of cases there were in each grade for some of the mental age levels. These data suggest that mental age is apparently a very important factor to consider in deciding at what point to begin or to complete the teaching of any phase of the long division process. This was also the conclusion of the Committee of Seven. The data suggest that this would not be at the same grade level for all pupils, or for all types of examples, indicating the need of considering carefully the means that may be used to adapt instruction to individual differences among pupils.

TABLE III

PERCENTS OF ERROR FOR PARTICULAR EXAMPLES BY MENTAL AGES FOR FOUR GRADES

Example	Grade	Mental Age in Months			
		120-131	132-143	144-155	156-167
86 ÷ 43	5B	12.5	5.3	0.0	0.0
	5A	0.0	5.4	16.7	0.0
	6B	0.0	10.3	2.6	2.9
	6A	25.0	6.3	5.9	0.0
785 ÷ 90	5B	33.3	36.8	18.8	16.7
	5A	34.5	37.8	33.3	12.5
	6B	29.4	30.8	15.4	11.8
	6A	37.5	43.8	35.3	11.5
90 ÷ 15	5B	47.9	34.2	13.6	0.0
	5A	37.9	24.3	55.6	0.0
	6B	33.3	20.5	12.8	11.8
	6A	50.0	18.8	14.7	15.4

In the second test the same difficulties in long division were tested. The number of quotient figures was increased to two and three, and zero difficulties in the quotient were included. The results of this test will be briefly summarized at this point. In general they support the findings that have been reported for examples of one-figure quotients.

Table IV contains an analysis of the difficulty of the various types of examples included in Test B. The data are so arranged by grades that comparisons can readily be made with the results for one-figure quotient examples given in Tables I and II.

A comparison of the results in Tables I and IV for all types of examples combined shows that the percents of error for examples with one-figure quotients tend to be slightly less than those for longer examples in all grades except Grade 5B. (See line (1) in each table)

A comparison of the percents of error in Tables I and IV for the two major types of examples, examples in which the apparent quotient is the true quotient and in which it is not (rows 2 and 3), shows that in both tables there is a very large difference in each grade in the difficulty of these two major types of examples. The types in which the apparent quotient must be corrected are by far the most difficult throughout. Moreover the general tendency is for examples with two or more quotient figures to be more difficult than those with only one quotient figure when examples containing similar skills are compared. Wide variations in the difficulty of the sub-types are also found in both tables. In both cases the percents of error on all sub-types where the apparent quotient must be corrected are very large, being as large as 70.4 percent in the type of example in which a combination of the difficulties included in the other types of examples is found (row 3g). The easiest type is the one in which one figure in the true two-figure quotient is one less than

TABLE IV

DIFFICULTY OF TYPES OF EXAMPLES IN LONG DIVISION FOR GRADES 5B TO 6A
(Two and Three Figure Quotients and Estimation Difficulties)

	5B	5A	6B	6A	No. of Examples
1. All types combined	42.3	48.1	31.8	33.7	
2. Apparent quotient is true quotient types combined	17.9	19.8	16.7	18.6	13
a. Range on examples of type	11.2–42.4	11.7–31.1	11.2–30.7	10.7–28.2	
3. Apparent quotient is not true quotient types combined	62.3	59.2	41.8	39.4	13
a. Is 1 less than apparent (one figure)	42.4	35.9	25.9	26.2	1
b. Is 1 less than apparent (both figures)	60.0	60.2	47.9	43.3	1
c. Type (96/9024) with or without remainder	68.4	63.6	48.7	52.5	1
d. Type (37/3034) with or without remainder	69.2	65.5	48.3	38.9	1
e. Teens (13/871) no remainder	57.6	50.5	36.7	30.1	1
f. Teens (18/1350) no remainder	56.4	46.6	25.2	24.3	1
g. Combination types (28/1896)	70.4	70.4	51.7	48.6	1
Range on examples of type	42.4–76.0	35.9–73.8	23.1–57.3	21.4–53.4	
Zero in Quotient Types					
4. All zero types combined	50.7	48.1	31.8	33.7	
a. End zero (2 figure quotient)	15.6	24.3	11.7	15.6	1
b. End zero and remainder	48.4	49.0	30.5	44.2	1
c. Middle zero (no remainder)	50.4	44.7	28.7	25.3	1
d. Middle zero and remainder	70.4	68.5	51.8	48.6	1
e. Double middle zero—no remainder	60.0	54.4	37.8	45.6	1
f. Double middle zero—remainder	65.6	56.3	47.6	42.7	1
g. Range on examples of type	14.4–72.8	24.3–69.0	9.3–51.8	14.6–55.3	
Number of cases	125	103	143	103	

the apparent quotient (row 3a). When both figures are one less than the apparent figures (see row 3b) there is a marked rise in the level of difficulty of the example. Examples in which the teens are the divisors (rows 3e and 3f) rank next in difficulty.

The types of examples with zeros in quotients are in general somewhat less difficult than the types in which the apparent quotient is not the true quotient and much more difficult than the types in which the apparent quotient is the true quotient. There is a considerable range in the difficulty of the sub-types which have zeros in the quotients. The type with a zero in the second place in the quotient (row 4a) is easiest. The type middle zero and a remainder (row 4d) is very difficult in all grades. In general the types in which there are remainders are the most difficult.

The retention test of the Committee of Seven contained, in addition to four one-figure quotient examples, three examples with two-figure quotients and three examples with three figure quotients. In seven of the ten examples the apparent quotients were the true quotients. Only

one zero difficulty was tested. It is therefore evident that this test was heavily weighted with relatively easy types of examples. This fact raises a further question as to the validity of the test as a measure of mastery of the total long division process.

To get at the relation between relative difficulty and mental age the data in Table V are presented. Here the results are arranged by mental age levels for the various types of examples so that the results can readily be compared with those of Table II which gives the data for examples of the various types with one-figure quotients.

Throughout the table for all types of examples the general tendency is for the percent of error to be smaller as the mental age increases. The percent of error for all types reaches 25 percent at the mental age level of 156-167 months (row 1). This is about two years later than that level of accuracy is reached for examples with one-figure quotients.

The types in which the apparent quotients are the true quotients are much easier than the other type at all mental age levels (row 2 and 3). The percent of error reaches 25 percent for the first type at the mental age level of 120-131 months, while for the second type this percent of error is not reached until the mental age level of 168 or more months, or about four years later. For sub-types 3b, 3c, 3d, and 3g the percent of error has not been reduced to 25 by the mental age of 168 or more months.

The data for zero-in-quotient types show that the index of 25 percent error, or 75 percent accuracy, on the whole is not reached before the mental age level of about 168 months, the same point as for the one-figure quotient examples in which the apparent quotient must be corrected to find the true quotient. There are wide variations among sub-types in the mental levels required to reduce the error to 25 percent; for example, type 4a requires a mental age level of 108 to 119 months, while to achieve this index of difficulty for several types apparently requires a mental age level of over 168 months, for instance, sub-types 4d, 4e, and 4f.

These wide differences in the difficulty of the two or three major types of examples and among the various sub-types raise the question as to the correctness of the position of the Committee of Seven that the process of long division should be taught to completion when pupils have the mental age of 12 years 6 months. Some of the easier steps can be taught a little earlier while others are so difficult that the problem should be seriously considered of reorganizing the program of teaching long division so that the work is more effectively distributed.

TABLE V

DIFFICULTY OF TYPES OF EXAMPLES IN LONG DIVISION BY MENTAL AGE GROUPS FOR GRADES 5B TO 6A COMBINED

	Mental Ages in Months						
	96–107	108–119	120–131	132–143	144–155	156–167	168 or more
1. All types	45.9	52.1	47.9	35.8	33.9	25.1	20.3
2. Apparent quotient is true quotient	51.9	28.6	24.3	14.9	17.6	14.6	11.4
3. Apparent quotient is not true quotient	78.8	73.9	67.4	51.9	45.9	32.7	27.7
a. Is one less (one figure)	75.0	50.0	47.8	38.3	28.1	13.7	12.5
b. Is one less (both figures	75.0	72.9	70.1	52.9	47.5	39.8	40.6
c. Type (96/9025)	100.0	72.9	73.9	60.5	57.2	39.8	38.7
d. Type (37/3034)	100.0	81.3	72.8	59.6	50.5	39.1	27.5
e. Tens by teens	100.0	72.9	62.0	47.1	38.5	26.0	17.6
f. Hundreds by teens	100.0	66.7	56.5	37.9	32.3	23.3	15.5
g. Combination types	100.0	79.2	78.8	60.5	58.3	37.7	35.6
4. Zeros in quotients	85.5	56.3	53.6	41.8	39.6	28.7	22.2
a. End zero—no remainder	67.5	25.0	22.8	15.0	17.4	7.6	8.7
b. End zero—remainder	75.0	47.9	52.8	43.8	38.9	31.6	23.7
c. Middle zero—no remainder	100.0	58.3	53.0	37.5	34.4	23.3	13.7
d. Middle zero—remainder	100.0	81.3	72.9	62.9	57.9	43.9	30.3
e. Double middle zero—no remainder	100.0	66.7	64.1	46.7	47.1	37.0	32.5
f. Double middle zero—remainder	100.0	70.8	69.6	52.5	51.2	37.0	37.5
Number of cases	4	24	92	120	121	73	40

We may summarize the major findings of this study and their implications for curriculum making as follows:

1. Long division is not a single general ability but apparently is a process that consists of a considerable variety of skills found in combination varying widely in difficulty.

2. It appears that the major type of example in which the apparent quotient is the true quotient is much easier for children of grades 5 and 6 than the major type in which the apparent quotient must be corrected to find the true quotient. This difference in difficulty is so great that the position can easily be defended that the two major types should not be taught at the same grade level. It is doubtful if this difference in difficulty can be reduced to any extent by possibly improved methods and materials of instruction.

3. This difference in difficulty between the two major types is further emphasized by the data concerning the mental age level necessary to achieve an index of less than 25 percent of error in the various types of examples. The findings for mental age level required to reach this standard for the major types classified according to kinds of quotients and difficulties based on the results of this study may be summarized as follows:

a. One-figure quotients
   (1) Apparent quotient is true quotient, mental age of 10 to 11 years.
   (2) Apparent quotient is not true quotient, mental age of 13 to 14 years.

b. Two-and-three-figure quotients
   (1) Apparent quotient is true quotient, mental age of 10 to 11 years.
   (2) Apparent quotient is not true quotient, mental age of 14 to 15 years.
   (3) Zeros in quotients, mental age of 13 to 14 years.

Here we find a difference of at least two years in the mental level required to achieve an index of less than 25 percent error on the two major types of examples for either one- or two-figure quotients. If a lower or a higher index than 25 percent error is considered to be more acceptable, the mental ages used as the basis of gradation would of course be modified accordingly.

4. A comparison of the figures for each major type reveals the marked differences there are among the various sub-types included in each group. This variability in difficulty should be taken into consideration in the gradation of the steps in teaching the process, and also in instructional procedure.

5. In general the mental level required to achieve an index of less than 25 percent error was the same for examples with either one or with two quotient figures when the apparent quotient is the true quotient. In the other type there was about a year's difference in the mental age level required to reach the index of less than 25 percent error in solving examples with one-figure and two-figure quotients. Examples with zeros in the quotients, some of which include these more difficult steps of correcting quotient figures, are about as difficult as examples with a one-figure quotient in which there must be a correction.

6. The facts included under points 3, 4, and 5 above indicate what may be an effective sequence of instruction. The need of carefully reviewing the recommendations of the Committee of Seven on this point is clearly indicated. It is very striking that the mental age level required to achieve a level of less than 25 percent of error when all types of examples with one-figure quotients are combined and the mental age level recommended in the report of the Committee of Seven are the same. However, in their report the Committee failed to differentiate between the level of difficulty of various kind of examples and did not include

many types in the test of achievement used as a basis of its recommendation. The present study supplies this essential information.

It should be pointed out in conclusion that the results of this study should not be considered as final. They represent conditions that result from practices in one well organized school system, where supervision is competent, and where carefully prepared instructional materials are used. It therefore seems reasonable to assume that the results of similar studies conducted elsewhere will not yield strikingly different results. However, it is quite possible that with even more carefully constructed teaching materials and with improved methods of instruction the general results in terms of accuracy of work man be somewhat better, but undoubtedly the general conclusion concerning the relative difficulty of the two major types and the various sub-types will not be changed to any considerable extent.

BORROWING IN SUBTRACTION

WILLIAM A. BROWNELL
Duke University

Editor's note: It is a commonly recognized fact that borrowing in subtraction is a difficult process for the average elementary school pupil. The author reports the results of a research that should facilitate learning in this field and illustrate an approach to yet other problems in this area.

IN an article by the writer in the May, 1938, issue of this journal passing reference was made to an investigation not then fully completed.[1] The results of this study, which related to borrowing in subtraction in the third grade, have since been reported at some length in a monograph.[2] Since, however, this monograph may not be accessible to the readers of this journal, the present review and summary of findings has been prepared.

THE PROBLEM

Teachers generally agree that children experience difficulty in mastering the process of borrowing in subtraction. To secure the correct answer for 467 — 235, little more is required of pupils than that they know the basic subtraction facts and that they subtract the pairs of digits in the right order. The example, 467 — 239, is, by contrast, much more difficult; here 9 cannot be taken from 7, and if the method of subtracting by decomposition is taught, a ten must be "borrowed."

At this point textbooks are not very helpful either to teacher or to pupil. If any explanation of the process is offered, the explanation commonly takes the form of a rule, and the children learn what they are to do, but not why or how. After an adequate amount of repetitive practice, they do get the trick, and they perform satisfactorily,—which is to say that they regularly get the correct answer in examples of this kind.

Some teachers of course have never been content with learning of such an unintelligent variety, and accordingly they have supplemented the textbook presentation in one way or another. One device which has been used consists in having the children change the form of subtraction examples as illustrated at the right (A). The tens' figure is reduced by 1, and the new tens' digit (5) is actually written on the work paper. Likewise, evidence of the borrowed ten is visibly shown by the inser-

(A)

$$\begin{array}{r} 5 \\ 4\ \not{6}\ {}^{1}7 \\ -\ 2\ 3\ 9 \\ \hline \end{array}$$

[1] Brownell, William A. "Two Kinds of Learning in Arithmetic," *Journal of Educational Research*, XXXI (May, 1938), p. 662, *et seq.*

[2] Brownell, William A., with the assistance of Kuehner, Kenneth G., and Rein, William C. *Learning as Re-organization: A Study in Third Grade Arithmetic.* Duke University Research Studies in Education, No. 3 (Durham, N. C.: Duke University Press, 1939), x + 87 + xx pp.

tion of a small 1 above and to the left of the ones' figure, thus making 17 instead of 7.

Teachers who employ this device do not usually advertise the fact, for they have been as much intimidated as have textbook writers by the oft-repeated dictum, "Other things being equal, form no habits which must later be broken." As a matter of fact, not one of the ten most recently published textbook series suggests this aid to children. One series makes use of the form shown at (B), and another outlines the form of thinking which is represented by (A) above; but no textbook author has dared to brave the opposition which would be aroused by recommending (A). So deeply have warnings against "crutches" entered our thinking that we hesitate to expose children to devices which, however helpful in initial learning they may appear to be, may remain permanently to impede advance to finished or expert performance.

(B)
5
4 6
− 2 3

Yet, on the surface of things, there seems to be some virtue in device (A). If children encounter difficulty in borrowing, it may be because they do not understand what they are doing and because, midway in the process, they lose their bearings and become confused. Device (A) conceivably could prevent error and confusion: It gives a tangible record of what must be done and of what has been done. Of course there is the possibility that the device does not even approach the real sources of children's difficulty in borrowing. There is the further possibility that even if it does, the ultimate consequences, in terms of slower rate of work, impaired accuracy, and infantile reliance upon a crutch, may be so bad that the device should not be taught.

The best way to find out whether device (A) is helpful or harmful seemed to be by experiment.

EXPERIMENTAL PROCEDURE

The procedure adopted called for four experimental groups of subjects. One group of subjects (the NC, or non-crutch section) were not taught the crutch at all,—as a matter of fact they never even saw it. The other three sections were taught the borrowing device, but under varying conditions which will be described below.

Four independent but parallel experiments were set up in certain third grades selected from all those available in the white schools of Charlotte, Greensboro, Raleigh, and Winston-Salem, North Carolina.

ach group of four teachers for each city chose their own coördinator d agreed upon a schedule which kept their four classes within two days each other in the program of instruction and measurement. The argin of difference in time among the four cities was approximately a eek, and this margin was maintained steadily throughout the experient.

All pupils in the four experimental groups (the class enrollment by ities ranged from 127 to 138, with a total of 590 pupils) were given an telligence test and an arithmetic Pretest to measure ability in the adtion and non-borrowing subtraction skills which had already been ught.

Following the Pretest, given in all sixteen classes between November and 24, 1938, the process of borrowing was introduced,—in the NC-ction by the usual rule-of-thumb procedure; in the three crutch sections explanations based upon use of the borrowing device. In all twelve rutch classes the teachers insisted that every pupil employ the device on ery subtraction example which involved borrowing. A sample daily sson was collected from all pupils about December 7, to ascertain the egree to which the device actually was being used. About December 17, about a month after borrowing had been introduced, Test I was given such a way as to furnish measures of (1) rate of work, (2) accuracy work, and (3) use or non-use of the crutch in borrowing examples.

At this time instruction for the three crutch sections (or the twelve rutch classes) was changed. In one crutch class in each city teachers ntinued to *require* use of the crutch in all daily work; hence this class designated the R-class, and the three such classes together, the R-ction. In another crutch class in each city (the D-class), further use the crutch was *denied* or forbidden. In the third crutch class (the O-class) the short form of borrowing without the crutch was shown, and e children were given an *option* as to whether they should adopt it or ould continue use of the crutch.

About two weeks later, daily lesson papers were again collected to termine the extent of crutch use; and two weeks later again, Test II as administered. This test was similar in design to Test I and was tended to yield comparable measures of rate, accuracy, and use or non-e of the borrowing device.

The administration of Test II on about January 14 concluded the periment proper. The children by this time had completed the text-ook chapter devoted to subtraction and had begun the study of multi-

plication. On about January 28, after two weeks during which very little practice on subtraction had been offered and during which nothing at all was said about the crutch, a third test was administered to all pupils. Test III consisted of three parts. In Part I the twelve subtraction examples all called for borrowing from the tens' figure alone (the only kind of borrowing as yet taught), and these examples were solved with no statement from the teacher as to whether the device should be used. Part 2 contained twelve similar examples, but the directions forbade use of the crutch to all children. Part 3 also consisted of twelve examples, but examples of unfamiliar types: the first six involved borrowing from the hundreds' figures, and the last six, borrowing from both the tens' and the hundreds' figures.

Test IV, comprising twenty-four subtraction examples, was given only to the pupils of R-, D-, and O-sections. Its purpose was purely to discover what had happened in the intervening two months and a half —whether the crutch children still relied upon the device or had abandoned it.

It should be added that throughout the experiment all pupils of all classes studied precisely the same arithmetical content and took all tests according to the same directions. In administering none of the tests except Part 2 of Test III, did the teachers say anything at all about use of the crutch. If crutch children inquired whether they should use the device, their questions were skillfully parried.

THE FINDINGS

It is impossible here to reproduce the nineteen tables and the seven figures in which the quantitative results of the experiment are reported in the monograph. Instead, only summary statements will be presented on the assumption that the interested reader will consult the fuller account in the monograph.

Soon after the raw data had been tabulated it became clear that the four independent experiments could not be treated separately. The four classes in each city had been carefully selected on the advice of superintendent and supervisor to insure rough equivalence of economic status of family and quality of pupil material. Nevertheless, large differences were found among the four classes of most of the cities, and it was necessary to make up composite sections of NC-, R-, D-, and O-children, to which the corresponding classes of each city contributed varying numbers of subjects. By this procedure the original population of 590 was reduced to 419, with about 100 in each experimental section.

These four composite sections ranged in mean I.Q. between 102.2 and 103.5 (the highest mean being in the NC-section) and in mean M.A. between 9-1.2 and 9-3.1. On the Pretest, the means for accuracy lay between 16.0 and 17.3, the poorest score being made by the 0-pupils. None of the critical ratios were reliable. As to rate of work, the medians (instead of means) were identical for all four sections.

The findings obtained from the series of tests and from other data faithfully recorded by teachers in logs may best be considered under three heads: *1.* Effect of the crutch on skill in borrowing; *2.* Permanence of use of the crutch, and *3.* Supplementary findings.

1. Effect of the crutch on skill in borrowing.

Accuracy. (a) On Test I all three crutch sections secured higher percents of correct answers than did the NC-children,—the means being 12.9 for Section R and D, 12.0 for O, and 10.9 for NC. Two of the differences in means were reliable, the critical ratio being 4.7 for R-NC and for D-NC. That for O- NC was 1.8.[3]

(b) On Test II, given after differences in instruction on the crutch had been instituted for Sections R, D, and O, the mean scores were: NC, 12.3; R, 13.6; D, 12.9; O, 11.8. In other words, two of the crutch sections were still superior to Section NC, though only the difference in means between R and NC was reliable. Apparently, then, the crutch actually assisted pupils to get correct answers from the start (Test I), and also (except for Section O) the advantage remained in favor of the crutch sections for some time (Test II).

(c) On Test III, given a full school month after the first instruction on borrowing, the mean scores on the three parts were as follows:

	Part 1	Part 2	Part 3
Section NC	10.3	9.9	7.3
Section R	10.5	10.0	8.1
Section D	10.6	10.3	8.0
Section O	9.4	8.9	5.3

Only the differences in means involving Section O were reliable. The small size of the other differences and the inconsistency of the advantage

[3] At the time of Test I, it will be recalled, all three crutch sections were being required to use the crutch; hence, theoretically the greater superiority of Sections R and D over NC as compared with that of Section O over NC is a bit difficult to explain. Or, it would be difficult to explain if other data did not indicate that the pupils of the O-section, in spite of the attempt to equate them with the pupils of other sections, were actually less capable in arithmetic.

therein among the sections probably together mean that by the time of Test III the NC-children had acquired the skills sufficiently well to secure as many correct answers as were obtained by the crutch children.

The uniform advantage in favor of the crutch children on Test I seemed to offer fertile ground for further analyses. Accordingly, all the records of the NC-children on this test were compared with the corresponding records of an equal number (121) of crutch children selected from all three crutch sections, in such a way that the NC-section and the new composite section of crutch children were equivalent in average I.Q., M.A., and rate and accuracy scores on the Pretest. Comparisons revealed that:

(d) The crutch children secured fewer incorrect answers on Test I not only for the borrowing examples but also for the non-borrowing examples. The ratios of errors for crutch and NC-children were 1.8 : 1 and 2.3 : 1, respectively, for the two types of example. There is a suggestion here that, confronted by the unintelligible task of borrowing in subtraction, the NC-children actually lost command over a skill (subtraction without borrowing) which they had supposedly mastered. This suggestion is borne out by other data to be presented shortly (paragraph g).

(e) The crutch helped not only children with low I.Q.'s (below 96) but also children with high I.Q.'s (above 113). Both high I.Q. and low I.Q. children who used the crutch were superior to high I.Q. and low I.Q. children who did not use it, and were superior both on borrowing and on non-borrowing examples.

(f) The same kind of finding resulted when children were divided on the basis of their accuracy scores on the Pretest: both the more able and the less able pupils profited from the use of the crutch.

(g) The NC-children failed to borrow 234 times when they should have borrowed, as compared with 11 times for the crutch children. Moreover, the NC-children made a total of 168 combination errors as compared with 98 for the crutch children. Their uncertainty about the process of borrowing apparently had spread so as to influence their control over facts they had learned and over a kind of subtraction which they had previously seemed to be able to deal with.

Rate. (h) On Test I the NC-children worked no more rapidly than did the crutch children, in spite of the fact that the latter children had been trained to write out their procedure. This rather surprising result seems to mean that *at this stage* in learning the thought processes of the

crutch children were as slow and cumbersome *without the crutch* as were those of the crutch children.

(i) On Test II Sections D and NC worked more rapidly than did Sections R and O. By the time of Test III, however, the R-children had pretty generally given up the crutch, and so their rate records were about the equal of those of Sections NC and D. The rate scores of Section O are inexplicable except on the basis of one or both of two assumptions: (1) that these children were generally inferior (as has been already suggested), and (2) that, not being told whether to discard or to retain the crutch, they were uncertain which to do and so tended to work both more slowly and more inaccurately.

With regard to rate, then, it can be said that use of the crutch did not penalize children who were taught it; they subtracted as rapidly as did the NC-children, both at first when the NC-children were fumbling about with the new process of borrowing and later when they (the crutch children) had discarded the device.

2. *Permanence of crutch*

It is of interest to know how firmly the crutch fastened itself upon its users. Devices like the borrowing device here studied are rather generally opposed in the belief that, once they have been taught, they can hardly be uprooted and that they will almost certainly be retained to the disadvantage of the users. The data collected with regard to permanence of the crutch are perhaps best considered in terms of the three sections,—R, D, and O.

(j) On Test I, 93.8 percent of the R-children used the crutch in at least one example. Also, the device appeared as a whole or as a reduced fragment in 86.6 percent of the examples. On Test II, the teachers in the interim having insisted upon use of the crutch in all written work, the corresponding percents were 87.4 and 88.6. On Part 1 of Test III the percents were 62.3 and 61.9, respectively; and on Test IV, the crutch was used in only 12.9 percent of the examples to which it was adapted.

(k) The D-children were forbidden use of the crutch after Test I. On Test I 99.1 percent of them used the device on 93.3 percent of the possible examples. On Test II these percents were 0.9 and 0.1, respectively, and on Part 1 of Test III, 0.0 percent and 0.0 percent. On Test IV the device was used with 1.3 percent of the examples.

(l) As previously pointed out, the records for the O-children are ambiguous. The percents of children using the device and the percents of examples on which the crutch was used were as follows: Test I, 93.7

and 87.3; Test II, 19.0 and 18.5; Test III (Part 1), 13.6 and 15.4. (In Test IV the crutch was used with 6.0 percent of the examples.

From these statements it appears clear that if use of the crutch is consistently opposed, the device quickly disappears or rapidly approaches the vanishing point (Section D); that if its use is required, many children will nevertheless insist on abandoning it (Section R); and that if given permission to discard the device, children will abandon it in steadily growing numbers (Section O). It also is clear from these figures that some children (a relatively small number) will continue to rely on the crutch long after it would seem to have lost any large amount of usefulness.

(m) No data were collected which reveal why certain children early gave up the crutch and why others retained it unduly long. No explanation could be found in term of C.A., M.A., I.Q. or initial ability in arithmetic. The explanation probably lies in traits which are essentially non-arithmetical, such as relative independence or docility, relative inventiveness and originality or want of the same, etc.

3. Supplementary findings

(n) Teachers, even those to whom the crutch was new, were unanimous with respect to ease of teaching it and with respect to its value in injecting understanding into a process which heretofore had been relatively unintelligible.

(o) A number of children who eventually surrendered the device clearly showed difficulty in so doing. Their daily papers and their test papers revealed various intermediate and shortened forms of the crutch before it disappeared completely. These successive alterations in the original form are, in the writer's opinion, best interpreted as further evidence of the contribution made by the device. That is to say, the fact that they could not give up the device all at once probably argues for their need for a kind of help which is usually withheld.

(p) The crutch, once it had been given up in connection with the familiar kinds of examples (in which a ten had to be borrowed), was not wholly gone. Rather, it remained available for cases of emergency, when it was revived and put to work. Certain illustrations of this function were found: (1) After a long period of non-practice, children regained their skill in borrowing by using the device a few times and then laying it aside again. (2) When they encountered more difficult kinds of borrowing (as in Parts 2 and 3 of Test II), they employed the device to help

in their unfamiliar tasks. (3) When they faced especially difficult applications of subtraction (as in problem solving and in "long" division), they fell back on the crutch as a means of holding steady one aspect of the situation which otherwise bristled with sources of confusion.

SIGNIFICANCE OF THE FINDINGS

The degree of significance which the reader will attach to the findings of this study will vary fundamentally with his conception of the purposes of arithmetic instruction and with his conception of the nature of the learning process in general. When the purposes of school arithmetic are viewed as the development of "efficient" skills,—mechanical tricks for manipulating figures quickly and correctly; and when learning is viewed as the process of giving the desired skills at once to children and then of leaving to practice the attainment of the desired proficiency,—when conceptions like these are held to, then the findings of the study are less than useful: they are distinctly subversive. They seem to offer support for a kind of instruction which cannot but interfere with the most "economical" learning on the part of school children.

But neither the purposes of arithmetic instruction nor the nature of the learning process need be conceived in these terms. When, to skill in computation as a purpose for arithmetic instruction, is added a second purpose, namely, that children shall see sense in what they learn, and when the learning process is viewed as consisting in a series of behavior stages which rise from beginners' first rather ineffectual but (to them) sensible efforts to the final stage of adult-like and intelligent performance,—under these conditions the findings of this study will be regarded as suggestive.

The borrowing device would be opposed on *a priori* grounds by those who think of arithmetic primarily in terms of skill and of learning primarily in terms of repetitive practice. Obviously, we do not want children forever to use the crutch in borrowing; its use, even initially, interferes with "economy" of effort; then, once it has been learned, it must be eradicated; and the whole process, it would be argued, is at best futile, and at worst harmful and wasteful. Confronted by the experimental evidence in this study, namely, that the crutch did not interfere with correctness of work (even added to it) nor with rapidity of work (except at later stages, when probably it should have been abandoned), the opponent of crutches would probably still insist that the device should not be taught. To teach the crutch is to take children off the straight course of

learning, to encourage them to develop a habit which they must then break,—and the evidence shows that not all the experimental children succeeded in breaking the habit.

To the writer at least this last kind of thinking is unacceptable; to him it seems stultifying and futile and even dangerous. There is implied in this thinking the notion that our knowledge about learning is now complete enough to determine by purely deductive means what is a good and what is a bad learning device or practice. Such complacency is hardly calculated to yield fertile suggestions and new ideas for the improvement of instruction nor yet to encourage research on learning as such. The test of worth in an educational practice is not whether it conforms, or can be made consistent with, what is now accepted as valid learning theory; rather, the test is, does it foster sound learning? In the present instance the borrowing device seemed to meet the test and accordingly it is recommended in the teaching of borrowing.

Still, one can hardly justify the space in this article (to say nothing of the space in the original monograph) if the only purpose is to sponsor a particular device in teaching subtraction. There is however further justification. The borrowing device should be regarded as typical of many other devices as yet not accepted and perhaps not even known which, when discovered and put to work, will make learning easier and more meaningful to children. This facilitation of learning is the basic responsibility of research workers in the psychology of the school subjects. If the meeting of this responsibility requires the abandonment of much that has passed as learning theory, then false learning theory should be cheerfully discarded in the assurance that a sounder theory will be found to take its place.

A NEW INCIDENCE OF LEARNING FOR THE FUNDAMENTALS OF ARITHMETIC[1]

GUY M. WILSON
Boston University

Editor's note: The subject of school grades has been of perennial interest. The author of this paper has raised a number of pertinent questions relative to grading practices in arithmetic and proposes that the percent of pupils attaining a perfect score be used instead of the commonly used mean or median.

THE word incidence, as used in the title to this paper means an index or indicator. In the past the incidence of learning may have been any of the following:

1. The passing grade—commonly 75 percent.
2. The norm—usually an average on a standardized test secured from a fairly wide testing program.
3. The percentile rank.
4. The grade attainment.

This is not a complete list. It does not, e.g., include the well-known McCall T Score, so mystifying to the average teacher. The passing grade doubtless occupies three-fourths to seven-eighths of the field the country over today. The other indexes listed above are a result of the application of statistics to the field of examining, grading, and promotion. In general, these various indexes are interchangeable; that is, any one can be translated into any of the others.

Types of Subject-Matter, Types of Teaching, Types of Testing? It is appropriate to ask what should determine the incidence of learning in any particular case. Are there types among which there may be different standards? Obviously there are, and this would easily be conceded in this group if applied to different subjects. Doubtless, anyone here would agree that any attempt to check or measure in the field of history would be very different from what it is if one were measuring handwriting or spelling or the fundamentals of Arithmetic.

If it is conceded that there are different types of subject matter as among different subjects, then doubtless there are some who would concede that there are different types of subject matter in almost any field and therefore in the field of arithmetic. Specifically, it is probable that everyone here would insist that a sixth- or eighth-grade pupil should know

[1] A paper read before the American Education Research Association at their annual meeting in Cleveland, February 25, 1939.

the 100 multiplication facts perfectly and should be able to use them automatically in examples. It is also probable that no one here would insist that a sixth-grade pupil should be required to know perfectly a table of Troy Weight or a table of Apothecary Weight, with real understanding of the meanings involved. These were at one time expectations in our schools, but studies of usage have modified this requirement. It is considered sufficient today that the child knows where to find these items if they are needed.

In other words, in the field of arithmetic there are varying types of subject matter. Some of these are essential drill and require perfect mastery because they constitute the much used processes in a subject that enters regularly into the transactions of daily life and business sufficiently to justify perfect mastery. The four fundamental processes of arithmetic are generally conceded to make up fully ninety percent of ordinary adult figuring. They constitute, therefore, the first consideration in a drill program for perfect mastery.

It is possible that the inclusion of useless items in tests in the past has had much to do with the acceptance of low averages as a basis of passing. If an arithmetic test included among its items 50 percent of relatively useless or non-drill material, as do the arithmetic sections of some achievement scales today, then a 50 percent mark should be passing. If 25 percent of the items are useless, then a 75 percent grade should be passing. But if all the useless or strictly non-drill items are omitted, and if the test is administered at the right grade level, then, may we ask, what should be the expected score? Does any one doubt the desirability of the perfect score in simple tests in addition and subtraction at the sixth-grade level? The making of poor tests in strictly drill subjects has, no doubt, been facilitated by the statistical fallacy, that all tests should contain some items so difficult that no pupil can make a perfect score. This fallacy should be recognized. Relatively useless items should be omitted from drill tests.

Just one word further on this point. There are different types of subjects. Three types are generally recognized, drill, problem, and appreciation. The fundamental purposes in these types are different. Should not the testing be different? Drill type testing in problem and appreciation subjects is detrimental. Yet few, if any, standardized tests have used other than drill types of testing.

It is well to recognize that some subjects contain no essential drill materials. It is also well to recognize that any subject that contains much

essential drill, also contains appreciation opportunities and may contain some problem opportunities. Now if we also recognize that non-drill items should not be tested with drill-type tests even in subjects where there is much drill, then we will have arrived at a point where we recognize at least the possibility of limiting drill type testing to essential drill items.

If now we apply proper criteria to drill, and retain only the items meeting these criteria, we arrive at the possibility of an entirely new standard.

What are proper drill criteria?

First, the material must be much used in daily living or it is not drill material.

Second, no attempt at mastery should be made until after clear understanding of meaning and adequate motivation.

Third, the load must be small enough that mastery is easily possible.

And the fourth item follows logically, viz., in essential drill only perfect scores are acceptable.

Now may we ask, instead of including improper items in a test and accepting low standards, would it not be better to omit the improper items? If we do this, is there not the possibility of new standards in line with the drill criteria just noted?

It will be well to consider this question somewhat in detail. In the first place it should be recognized that a passing grade somewhere around 75 percent still prevails in most schools. In the second place it will be worth while to consider the meaning of norm or grade attainment or percentile rank, and whether or not these can be expressed in terms of a percentage similar to the old passing grade. In the third place it will be worth while to analyze the meaning of an average, norm or grade attainment by distributing grades and noting exactly what is meant by a norm or a percentage grade particularly in terms of the distribution of pupils on a complete scale running from one hundred percent down to zero.

In this presence it may be assumed that the first two items are fully understood. But careful study of the third item, viz., the meaning in terms of distributed scores, of an average or norm, convinces me that in the past we have failed to understand the significance of these measures which we have used so freely.

The Significance of an Average in Terms of Distributed Scores. The discussion of drill criteria above indicates that the important consideration in strictly drill subject-matter, is the percentage of children making

no errors, or securing perfect scores. What does a particular average mean in terms of perfect scores?

Here are returns from 31 different groups at the sixth-grade level on the Wilson Addition Process Test, in terms of the mean as average and the corresponding percentage of perfect scores. The results are rounded off to avoid decimals.

Mean (Addition Process Test, 6th Grade)	Percentage of Perfect Scores (25 items all correct, score 100%)	Mean (Addition Process Test, 6th Grade)	Percentage of Perfect Scores (25 items all correct, score 100%)
99 (99.34)	89	88	13
99 (98.70)	74	88	10
96 (95.60)	48	88	9
96	25	87	15
96	19	87	14
92	27	87	12
92	17	84	12
90	19	82	0
90	14	80	24
89	17	80	0
88	18	76	3
88	18	76	0
88	17	72	7
88	16	72	0
88	14	52	0
88	14		

This small exhibit of means and corresponding percentage of perfect scores will be clearer from a study of the following distribution tables. It is evident, however, that a high average such as 96, may signifiy relatively few perfect scores, and that a mean as high as 80 may indicate no perfect scores, i.e., every pupil making errors on this simple test.

The Wilson Addition Process Test. As a basis for continuing the discussion let us direct attention to the test used as a basis for the above tabulation. The Wilson Process Test in addition is as simple as adult usage. The most difficult example has only six two-place addends. No addition fact is needed above 39 + 9. Every item is reasonable. Some items are merely primary facts. Here, then, is a test so simple that no one will deny that perfect mastery is a reasonable expectation in grade six or above. Additions in non-drill fractions, denominate numbers, and decimals do not occur.

Furthermore the test consists of a total of twenty-five examples. All of the one hundred primary facts are covered and over one hundred of the upper decade facts are covered. All of the common addition process difficulties appear. Obviously this simple test is adequate for a reasonable checking of pupil mastery, or a diagnosis of pupil difficulties.

What Time Allowance on a Simple Drill Test in Addition? The next question is how much time should be allowed for the test. Should pupils

be permitted to complete the test in order that we may see what they do, that is, should the time be permitted to run on as needed in order to get also the diagnostic value out of the time taken for the test? If the purpose is to aid in teaching, evidently the pupil must be permitted to complete the test, especially in consideration of its simplicity, and the fact that it is being given in one of the upper grades. In a sixth grade in a sizable New Hampshire city, the time ran from four minutes to twenty minutes. In the sixth grade in fifteen cities in the Boston area the time ran from one minute to twenty-five minutes. A very small percentage did not complete the work in twenty-five minutes, nearly two percent. The time taken helps in diagnosis of pupil needs.

What Score? From the standpoint of the average score it may be noted that the New Hampshire city referred to above had a median score of eighty in the sixth grade and a mean score of 78.8. This score compares favorably with the usual 75 percent passing. It is, however, a little low in comparison with other cities. For fifteen cities in the Boston area the means on this test run as follows: fourth grade, 65.36; fifth grade, 80.88; sixth grade, 87.96; seventh grade, 88.65; eighth grade, 89.83.

The Meaning of an Average. These averages seem unusually satisfactory, yet they are not satisfactory, as a careful spread of data will easily show. The New Hampshire city with a median of eighty in the sixth grade showed six perfect scores out of 170—less than four percent perfect. The fifteen towns and cities in the Boston area showed from nine to nineteen percent perfect scores at the sixth grade level. That is, these were the results in fourteen cities. In one city; city 0, there were 89 percent of the children showing perfect scores and this again suggests the possibility of entirely new standards. Using, as a basis for thinking, the most favored situations among my coöperators for the last four or five years, we may note mean scores for grades four to eight on the Wilson Addition Process Test as follows: grade four, 98.00; grade five, 99.10; grade six, 99.34; grade seven, 99.58; grade eight, 99.86; and still the question is what do these scores mean. In the case of the sixth grade mean of 99.34, we have the one city, City O, noted above, among the fifteen in the Boston area where the perfect scores equalled 89 percent. The other eleven percent of scores for City O, were distributed as follows: at 96, 7 percent; at 92, 2 percent; at 88, 2 percent.

In other words a mean score of 99.34 in this case means that there were still eleven percent of children not perfect.

If a mean score of 99.34 can mean that there were still eleven percent

of the children not attaining perfect scores, what does a mean score of 75 or 80, or 85 signify? We have noted that the mean of 78.08 in a New Hampshire city meant that about four percent of the children had perfect scores. Furthermore in that city about eight percent had a score of 96, about twelve percent a score of 92, and the scores continued down to 56 where four percent of the pupils registered.

Taking one of the more favorable of the fifteen cities in the Wilson W.P.A. study in the Boston area[2] (or rather fourteen cities if we omit City O, the exception to the rule as noted above), City E had a median of 92 percent. The mean was 89.42. These scores appear to be very favorable. They are, as compared with the 75 percent average. Distributing the scores, however, we note that: 17 percent fell at 100, 22 percent fell at 96, 18 percent fell at 92, 15 percent fell at 88, 9 percent fell at 84, 8 percent fell at 80, and on down through 76, 72, 68, 64, 56, 48, 36, and 24 to zero. In other words a mean of 89.42 indicated only 17 percent perfect scores.

Taking other averages, we may note that a mean of 87.93 shows 9 percent perfect scores. Another mean of 89.91 shows 19 percent perfect scores. Time and again the scores run down as low as 56. Some data with similar implications are presented in an article scheduled for the April number of the *Mathematics Teacher*.[3]

Distributing the Scores. The obvious point to the above is that a mean score or any kind of an average or norm taken alone requires very careful interpretation. The most obvious method of interpretation is to spread the scores along the entire scale and if we are interested in fundamental diagnosis it evidently means that the distribution should include for each child, not only score but time. This two-way distribution based upon score and time will facilitate an interpretation and will help decidedly toward the conclusion that in a test so simple as the one under consideration, and with a time that is not limited, the pupils at the sixth-grade level and above should show perfect scores or their work is not satisfactory.

From among hundreds of tables showing two-way distributions for score and time on the Wilson Addition Process Test, a few typical ones are here presented.

Table I shows a two-way distribution, score and time, on a room of 26 sixth-grade children in one of the better school systems in the Boston

[2] See *Yearbook of the American Educational Research Association*, 1937, p. 274.

[3] Wilson, Guy M., and Hanley, Gertrude L. "For Basic Drill in Arithmetic, What Norm or Average Is Satisfactory?" *The Mathematics Teacher*, April, 1939.

area. It will be noticed that there are no perfect scores although the median score is 72. This is one of the poorer of the six sixth-grade rooms in the building. In Table II, it will be noticed that eighteen percent of the scores fall at one hundred, ten percent at ninety-six, and that the low point is seventy-two. This is one of the better sixth grades in this building. The median is 88. It shows results that are typical for the better traditional schools throughout the country.[4]

Careful study of Tables I and II will help one to realize that we have

TABLE I

DISTRIBUTION OF SCORE AND TIME ON A. P. TEST MADE ON INITIAL TEST BY ROOM N. SIXTH GRADE, OCTOBER, 1938

Score	Time in Minutes														Total	Percent
	4	5	6	7	8	9	10	11	12	13	14	15	16	23		
100															0	
96			1			1									2	7.68
92															0	
88	1				1	1		1							4	15.36
84							1								1	3.84
80						2			1	1					4	15.36
76				1						1					2	7.68
72						2				1					3	11.52
68						1		1				1			3	11.52
64											1				1	3.84
60									1					1	2	7.68
56										1			1		2	7.68
52								1							1	3.84
48				1											1	3.84
Totals	1		1	2	1	7	1	3	2	4	1	1	1	1	26	99.84
Median score			72													
Mean of scores			73.7													
% Perfect			00.													

TABLE II

DISTRIBUTION OF SCORES AND TIME ON A. P. TEST MADE ON INITIAL TEST BY ROOM W. SIXTH GRADE, OCTOBER, 1938

Score	Time in Minutes											Total	Percent
	5	6	7	8	9	10	11	12	13	14	15		
100		1		1	1	1	1					5	18.05
96			2			1						3	10.83
92			1									1	3.61
88			1	1		2	2					6	21.66
84			1		1				1			3	10.83
80							2		1			3	10.83
76			2			1		1				4	14.44
72							1		1			2	7.22
Totals		1	7	2	2	5	6	1	3			27	97.47
Median Score		88											
Mean of Scores		86.9											
% Perfect		18.05											

permitted children even in the most fundamental drill material to go forward with scarcely the semblance of mastery. Such results in addition would not be tolerated in any business. For all practical purposes, there-

[4] Tables I and II are due to the coöperation of Florence Ridlon, Graduate student.

fore, they are worthless. The time spent has been wasted, probably worse than wasted since it has built up aversions that keep these children permanently insulated from arithmetic.

Two other tables are inserted here for comparison.[5] These come from another school. Table III shows the results on the Wilson Addition Process Test in a sixth grade in September. No child made a perfect score, although the median is 84, and the mean 82.4. In the meantime, however, the teacher had taken courses in the teaching of arithmetic. She had developed the concept of better work, and determined to remedy the situation.

Table IV shows the results of using the same test over, (duplicate form) three months later, (December), with the same sixth grade. Table IV thus shows the results on the same test after three months of special work. The mean rises from 82.4 to 95.6. Approximately 50 percent of the scores are perfect. Table V shows results of March testing of Short Division by another specially trained teacher.[6]

Under all the circumstances, what should be the incidence of learning in the fundamentals of arithmetic that will be most helpful?

The incidence of learning that is proposed is the percentage of the pupils with perfect scores, not the mean or median or some norm or grade attainment based upon one of these, but the percentage of perfect scores even though the mean may run as high as 99.34. This high average as

TABLE III

DISTRIBUTION OF SCORES AND TIME ON THE A. P. TEST
SIXTH GRADE, SEPTEMBER 7, 1938

Scores	Time in Minutes														Totals	Percent
	9	10	11	12	13	14	15	16	18	19	20	27	28	71		
100																
96				1	1				1						3	9.0
92			2	1									1	1	5	15.0
88		1			2	1					1				5	13.0
84	2	1		3	1					1					8	23.0
80				1			1								2	6.0
76		1		1	1			1		1		1			6	17.0
72	1			1											2	6.0
68						1	1								2	6.0
64																
60																
56						1									1	3.0
Totals	3	3	2	8	5	3	2	1	1	2	1	1	1	1	34	100.0

Median Score 84
Mean of Score 82.4
% Perfect 00.

[5] Kindness of Mary Kett, Needham, Massachusetts.
[6] Kindness of L. K. Houghton, Jr., Wayland, Massachusetts.

TABLE IV

DISTRIBUTION OF SCORES AND TIME ON THE A. P. TEST
SIXTH GRADE, DECEMBER 9, 1938

Scores	Time in Minutes												Totals	Percent
	5	6	7	8	9	10	11	12	13	14	15	16		
100	2	3	2	2	1	1	2	2				1	16	49
96	1	1	1	2	2								7	21
92	1			2			1						4	12
88			1	2									3	9
84			1	1		1							3	9
Totals	4	4	5	9	3	2	3	2				1	33	100
Median score	96													
Mean of scores	95.6													
% Perfect	48.4													

TABLE V

DISTRIBUTION OF SCORES AND TIME ON THE S. D. P. TEST
EIGHTH GRADE—A RE-TEST, MARCH 3

Scores	Time in Minutes							Totals	Percent
	2	3	4	5	6	7	8	9		
100	1	4	5	3	5	2	1	1	22	95.6
96		1							1	4.4
Totals	1	5	5	3	5	2	1	1	23	100.0
Median score	100									
Mean of scores	99.83									
% Perfect	95.6									

noted above, still left 11 percent of the pupils with scores of less than 100. The percentage of scores not reaching 100 percent, becomes the significant index. It indicates the task ahead. A mean of 99.34 is not satisfactory if it indicates 11 percent of the children without perfect scores. Nor is a mean of 99.99 satisfactory. The only acceptable mean for a group of upper grade children in a simple test in a fundamental process is the mean of one hundred.

There should be no hesitancy in accepting the percentage of scores at 100 as the correct incidence of learning in simple drill material such as is involved in the fundamentals of arithmetic in whole numbers. If the test is right, if the curriculum is right, and if the teaching is right, the perfect score in simple useful drill is possible for any normal child. The use of the proposed incidence is more discriminating if comparisons of classes or cities are wanted; it leads to more fruitful diagnosis; and it offers a challenge to the highest grade of teaching.

APPLICATIONS OF ANALYSIS OF VARIANCE TO EDUCATIONAL PROBLEMS

JACK W. DUNLAP
University of Rochester

Editor's note: Persons who secured their basic training in statistics some years ago may or may not have kept up to date. The author presents a discussion of an important development in this area.

IN order that there be no misapprehension as to the content of this paper, let me state now that nothing new will be presented, no knowledge of higher mathematics assumed, or complex formulas developed. This paper is designed to serve a far humbler purpose in the general scheme of educational and psychological research.

The development of statistical techniques and the growth of the number of applications during the past five decades might be regarded as something similar to a geometric progression. In view of this tremendous development and growth, it is not at all astounding that valuable techniques should be overlooked. It is the purpose of this paper to remind you of one of such technique, the analysis of variance, in the hope that you will consider the method in the light of your own specific problems. There can be no question that analysis of variance is an extremely powerful method, but it remains for research workers to make applications to their problems for the method to spring into prominence. A few simple applications will be presented to titillate your interest in this tool.

Experiments are often designed to test the "null hypothesis"; that is, the hypothesis that there is no population difference in the varieties under consideration. Essentially this is what is done when a difference between two means is determined, the standard error of the difference computed, and the difference tested to see whether it is significantly different from zero. However, it should be remembered that it is never possible to prove a null hypothesis.

Analysis of variance is a technique developed by R. A. Fisher for segregating from comparable groups of data the variability assignable to different known sources. Actually it is a method for arranging arithmetical facts so as to isolate into two or more parts the variation traceable to each specified source. This segregation of the variation attributable to specific causes makes it possible to apply either Fisher's z-test, or the more readily applied F-test of Snedecor for detecting significant differences among the sub-groups within a set of data. Another important

advantage of so arranging the data is the provision for the estimation of experimental error independent of the variabilities whose sources are known or assignable.

Whatever the use, the arithmetic involved is simple. In its simplest form, the method involves three quantities, the total sum of the squares, the sum of the squares within classes, and the sum of the squares between classes. The total sum of the squares consists of the squared deviations of each individual about the general mean. The sum of squares within classes consists of the squared deviations of the individuals from their group means. The sum of the squares between classes is the difference between these two. In practice, for computational purposes only, one reverses this order, first securing the total sum of squares. Then one determines the sum of the squares between means of classes around the general mean, weighting each mean by the class frequency.

The discovery by Fisher of the distribution of ratios of these sums of squares makes possible, the test, for significance of the dispersion among class means with the dispersion to be expected on account of random sampling. A more convenient form of Fisher's z-test of these ratios is Snedecor's modification, namely, $F = \frac{\text{Larger mean square}}{\text{Smaller mean square}}$. Efficient methods in the design of experiments and the use of techniques such as the analysis of variance should bring about substantial reduction in the time and cost of experiments.

The method of analysis of variance rests on two points: (a) the additive character of sums of squares, and (b) the assumption that the variate under consideration is normally distributed in the population sampled. In practice, considerable leeway is allowed in departure from normality, but obviously this imposes restrictions on the interpretation of the results. With these preliminary remarks as a stage, or perhaps as a verbal barrage to cover the paucity of examples, consider the following applications of analysis of variance:

The limitation of the services of the statistician in a school system to studies of pupil accounting, educational status, and test construction has always seemed to the writer an undue prescription of the statistician's services. Bureaus set up for testing school supplies, school lunches, and so forth, should avail themselves of the statistician's tools in evaluating the efficiency of materials and methods. Consider the following hypothetical example: Complaint had been registered by the teachers in city A concerning the chalk used in the classrooms to the

effect that it was too hard and contained an undue amount of grit. The Bureau of Standards found variations in hardness among the commercial chalks of several companies, but the question remained as to whether these variations might be attributable to sampling errors. Five sticks selected at random from each of a dozen boxes of a given brand purchased in a variety of stores over a period of time furnished samples of each brand. Seven brands were under consideration, the schematic analysis of which is given in Table I.

TABLE I

ANALYSIS OF BRANDS OF CHALK FOR HARDNESS

Source of Variation	Degrees of Freedom	Mean Square	
		Situation A	Situation B
Total	419	120	120
Brands	6	60	60
Boxes of same brand	77	40	20
Crayons from same box	336	20	20

Two possible situations or sets of results are indicated in Table I. Both of these are fictitious, and it should be pointed out that, in practice, one does not get such convenient even numbers.

A few comments on the various degrees of freedom given in the table might be helpful. Since there are seven brands, twelve boxes of each brand, and five samples taken from each box, the total number of crayons is 420. One degree of freedom is lost due to the use of the mean in getting the sum of squares, so that there are only 419 degrees of freedom. In the study of the seven brands, again one degree is lost, so that only six remain. In the study of boxes within brands, one degree is lost for each brand, so that there are 7×11, or 77 degrees of freedom here. In the study of crayons from the same box, we have only 4 degrees of freedom in each box, so that the total is $4 \times 12 \times 7$ or 336 degrees.

Under situation A, the boxes of the same brand do not differ significantly, since 60/40 give an F value of 1.5. The boxes in the same brand, however, differ significantly indicating a lack of quality control in manufacture as revealed by the F ratio of 40/20 or 2.00. The 1 percent point here is 1.84. The question now arises as to whether this variation holds for all seven brands. This requires seven analyses to be made, each as indicated in Table II.

TABLE II

ANALYSIS OF QUALITY VARIATION WITHIN A SPECIFIC BRAND

Source of Variation	Degrees of Freedom
Total	59
Boxes within a brand	11
Crayons within a box	48

If, however, the mean squares had been as indicated under the second situation, B, then uniform quality would have been indicated for each brand, but the brands would have been judged to differ significantly among themselves.

The old question of the relationship between high school preparation and college success has been reopened recently by Dressel (1). To what extent do high schools differ in the grades assigned to students who go to college, and are these diffrences reflected in their college marks? Dressel's study is of two-fold interest, first, for the topic, and second, for his use of the analysis of variance technique. He studied the records of 810 students distributed among 15 high schools, for whom both high school records and first year college averages were available. In Table III is presented the basic data from Dressel's study.

TABLE III

ANALYSIS OF HIGH SCHOOL AND COLLEGE GRADES

Source of Variation	Degrees of Freedom	Sum of Squares		Mean Square	
		High School	College	High School	College
Total	809	302.41	512.30	.37	.63
Between means of high schools	14	28.81	19.51	2.06	1.39
Within high schools	795	273.60	492.79	.34	.62

The total sum of the squares for high-school grades is 302.41, but this can be broken down into two parts, that attributable to the difference between means of high schools, and that due to individual variation within high schools. Now it might be postulated that there are no differences between high schools. If this is the case, then the mean squares representing either the variation between the means of high schools or variation within high schools could be used as an estimate of the variance of school grades. If the null hypothesis (that either of the values may be used as an estimate of school grades) holds, then these values should not differ significantly. Now, dividing the variance between high-school means by the variance within high schools gives 2.06/0.34 or 6.01. Entering Snedecor's table with 14 and 795 degrees of freedom respec-

tively, it is found that the ratio of 2.20 corresponds to the 1 percent point; therefore, the differences between high school averages are highly significant. Similarly, an analysis of the college grades for the same groups of students gives a ratio of 1.39/0.62 or 2.25. This is interpreted as showing that the high-school groups differ significantly in the grades received in college. The mean square among high-school grades is considerably larger than the mean square among means of college grades for the same high-school groups. This indicates that, so far as the college is concerned, there is not nearly as much difference between the groups from various high schools as the high-school grades would lead one to expect. It is not possible to compare directly the values of 2.06 and 1.39 since there is correlation between the college and high-school grades. This brings up the problem of co-variance, a natural and highly important extension of the problem, but not treated here because of lack of time.

Another type of problem that continually arises to plague the investigator is illustrated by the following: Most school populations are heterogeneous with regard to race, and the investigator is forced to consider whether or not the ethnological factors will disturb his results. Another ever present question is whether sex differences will invalidate the results. Dunlap (2) in a survey of the public school children in Honolulu with regard to achievement as measured by the arithmetic reasoning test of the old Stanford Achievement Test was confronted with a series of such questions.

The first question, namely, Is the difference between the means of races sufficient to disturb the results? might be inverted and the postulate made that there is no real difference between the means of the various racial groups. The computation of the various sums of squares was done in the usual manner and the results are shown in Table IV.

The mean square within sex-race subclasses may be taken as an esti-

TABLE IV

ANALYSIS OF RACE VARIATION IN ARITHMETIC REASONING

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Square
Total	3758	2,430,025	647.48
Between means of races	5	169,400	33,880.00
Between means of sexes in race, one degree of freedom in each race	5	171,010	34,202.00
Within sex-race sub-classes	3748	2,089,615	557.52

mate of experimental error, so that the mean square between means of races is divided by that within races. This ratio, 38,880/557.53 is 69.73. No table is needed to interpret this, for it is obvious that the means of the races vary significantly. It can be shown in a similar fashion that the sex differences within races are significant. The wide discrepancies in the means of races are not attributable to chance and mingling of such groups might well distort correlational studies of such data.

A more complicated problem that might be presented to the hypothetical Bureau of Standards mentioned above is exemplified by the work of Duncan (3), quoted in Snedecor's admirable volume (4). Duncan's problem was to determine the merits of lard as a cooking fat. Each fat was used in making four mixes of dough, and from each mix, forty cookies were baked. The individual cookies were then tested for breaking strength. The problem was to test the technique of mixing, baking, and breaking the cookies, to make a valid estimate of experimental error and to determine whether the fats differed significantly in respect to breaking strength of the cookies made from them. Snedecor presents the following analysis, Table V.

TABLE V

ANALYSIS OF TENSILE STRENGTH OF COOKIES MADE FROM DIFFERENT FATS

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Squares
Total	479	30,399	—
Within mixes	468	3,269	6.99
Between means of fats	2	26,573	13,286.0
Between means of mixes within fats	9	557	61.9

It will be noted that the sum of the squares, 27,130, between the means of the twelve mixes is not entered in the table, since it is not needed except to determine by subtraction the variance ascribable to "within mixes." This problem is comparable to the chalk problem but set up slightly different. Note that these twelve means account for eleven degrees of freedom. There are three fats; therefore, there are only two degrees of freedom between means of fats. Since each mix mean is measured from its own fat mean, there are only three degrees for each fat, or a total of nine between the means of mixes within fats. The sum of the degrees of freedom between means of fats, two, and between means of mixes within fats, nine, gives a total of eleven, which checks with the degrees of freedom between the twelve mix means.

There are two striking features in the above analysis,—first the gre difference between the mean square between mixes within fats and th within mixes. This indicates a failure in technique, for the four bakin from each fat differ significantly from each other. The second feature the tremendous mean square between fats, which, when compared wi the value 61.9 indicates that the fats are significantly different.

Dr. Grace Rubin-Rabson (5) in a recent study of the psychology memorizing music has used analysis of variance effectively in the interpr tation of her data. The nature of the problem was such that the numbe involved were of necessity small, only nine cases. The nine subjects we divided into three groups, roughly equivalent as to training and exper ence. Nine compositions were employed, each group learning three com positions by each of the three methods. The methods, the compositions and the order of learning were rotated so that every combination method and composition was presented in every order. In order to d this, it was necessary to arrange each of the three groups into a Lati Greek square construction, and the designs for the three groups i turn constituted a large square, so that the varying capacities of th subjects operated equally as did the varying difficulties of the compos tions. The procedure involved 81 learnings, 27 by each of the thre methods.

TABLE VI

SCHEMATIC ARRANGEMENT OF ANALYSIS OF MUSIC DATA

Source of Variation	Degrees of Freedom
Total	80
Subjects	8
Methods	2
Compositions	8
Order	2
Residual error	60

The following example is hypothetical. It is proposed to investigat the relative efficiency of teaching science by three different method (1) regular classroom instruction, (2) classroom instruction supple mented by radio talks and questions formulated by various classes an swered over the radio, and (3) classroom instruction supplemented b "canned" lectures, the "canned" lectures to be identical with the radi talks, but no discussion other than with the classroom instructor. Th problem in its simplest form can be thought of as consisting of two experi mental groups and one control group. Thus only three instructors would

be necessary. The results, however, probably would be biased by variations in teacher efficiency and interest. This may be partially controlled by having each teacher give instruction under each experimental condition. A further difficulty arises due to variation of student ability from class to class. This difficulty gives rise to the problem of experimental design, which merits extensive comment, but must be omitted for lack of time.

The actual scheme of analysis of variance for such a study might well be set up as in Table VII.

TABLE VII

SCHEMATIC ANALYSIS OF DATA ON VALUE OF RADIO TEACHING

Source of Variation	Degrees of Freedom
Total	N-1
Within Classes	N-9
Between means of teacher groups	2
Between means of method groups	2
Interaction or remainder	4

An analysis such as the above provides tests for most of the basic questions arising in analyzing the results.

Time does not allow further examples, but in closing the author should like to call attention to an article by Lev (6) entitled "Evaluation of Test Items by the Method of Analysis of Variance." This is another unexplored field for the application of this method.

In conclusion, let us summarize the above. The method of analysis of variance is not new; it is an arithmetic technique; in the case of two variables it gives exactly the same results as Fisher's t-test for the difference between means; nevertheless it is a useful and valuable tool for the research worker. Examples have been given to illustrate its usefulness (1) in the analysis of causal relationships of school grades, (2) in determining presence or absence of heterogeneity among subjects, (3) in the testing of differences in products, (4) in establishing quality control both in manufacturing and experimental conditions, and (5) in experimental problems of learning and test construction.

The omission of the analysis of co-variance, a logical extension of the method of analysis of variance has been deliberate, since the purpose of this paper is to emphasize areas in which the method is applicable, not the extensions of the method. Likewise, computation has been omitted, since it is simply and adequately described in Snedecor's two volumes (4, 7).

BIBLIOGRAPHY

1. Dressel, Paul L. "The Effect of High School on College Grades," *Journal of Educational Psychology*, XXX, 1939. In press.
2. Dunlap, Jack W. "Race Differences in the Organization of Numerical and Verbal Abilities," *Archives of Psychology*, 1931, No. 124.
3. Duncan, Helen J. "A Comparison of the Shortening Power of Lards with other Fats and Oils in plain eggless Cookies and in Pastry." A Thesis submitted for the degree Master of Science, Iowa State College, 1933.
4. Snedecor, George W. *Calculation and Interpretation of Analysis of Variance and Covariance*, Ames, Iowa: Collegiate Press, 1934.
5. Rubin-Rabson, Grace, "Studies in the Psychology of Memorizing Piano Music, A Comparison of the Unilateral and the Coördinated Approaches," *Journal of Educational Psychology*, XXX (1939), pp. 321-345.
6. Lev, Joseph. "Evaluation of Test Items by the Method of Analysis of Variance," *Journal of Educational Psychology*, XXIX (1938), pp. 623-630.
7. Snedecor, George W. *Statistical Method*. Ames, Iowa: Collegiate Press, 1938.

THE INTELLECTUAL GROWTH OF COLLEGE STUDENTS

CHARLES D. FLORY
Lawrence College

Editor's note: At what age does one reach intellectual maturity? There has been some considerable difference of opinion about this matter. The author presents data to show that college students at Lawrence College, Wisconsin, make statistically significant gains in intellectual maturity during the years that they are in college.

THE end point in intellectual development has been a controversial issue for several decades. Early investigators were so absorbed in finding adequate testing instruments that they agreed quite arbitrarily upon the age at which individuals reach intellectual maturity. When experimental data began to reveal conflicting results the maximum mental age was revised up or down to meet the latest body of evidence. The effort of psychologists to adjust their thinking to the experimental findings caused the upper limit of intellectual growth to be pushed all over the "teen ages," with age sixteen quite widely accepted for test purposes. Recent evidence[1] has shown that intellectual growth continues through the "teen ages" and likely into the twenties when there is stimulation appropriate to the maturity level of the individuals concerned.

Rogers[2] was one of the first investigators to show that college students actually improve in intellectual ability as measured by present-day intelligence tests. Re-tests were made on 56 seniors, 35 juniors, and 55 sophomores at Bryn Mawr College. These re-test results were compared with the freshman scores on the Thorndike Intelligence Test. Analysis of these data revealed that the amount of growth was related to the length of time the student had spent in the institution. Seniors had gained more than juniors, and juniors had gained more than sophomores. Rogers concluded that although the gains in intellectual ability after age eighteen were small they were significant.

[1] Freeman, Frank N., and Flory, Charles D. "Growth in Intellectual Ability as Measured by Repeated Tests," *Monographs of the Society for Research in Child Development*, II, No. 2 (1937), pp. xi + 116.

Hartson, L. D. "Does College Training Influence Test Intelligence," *Journal of Educational Psychology*, XXVII (October, 1936), pp. 481-91.

McConnell, T. R. "Change in Scores on the Psychological Examination of the American Council on Education from Freshman to Senior Year," *Journal of Educational Psychology*, XXV (January, 1934), pp. 66-69.

[2] Rogers, Agnes I. "The Growth of Intelligence at the College Level," *School and Society*, XXXI (May 24, 1930), pp. 693-99.

Wright[3] found from re-test data at the University of Pennsylvania that college students had not reached intellectual maturity. He concluded from his study, "Though we may not expect the complex mental processes to increase indefinitely, there is no evidence for an approaching plateau at age twenty-three."

McConnell[4] discovered significant improvement in the test scores of Cornell College students after four years of college training. The scores for 70 seniors showed a gain of more than 40 points on the American Council Tests when compared with the freshman scores. A critical ratio of 18 plus indicates a highly significant improvement in intellectual ability. The women in the Cornell study gained nearly 12 points more than the men but the sex difference was not statistically significant, (a critical ratio of 1.73). McConnell did not report the number of men and women in his study. He found a correlation of .83 ± .02 between freshman scores and senior scores.

Hartson[5] studied the effect of college training on the intelligence of 394 Oberlin seniors. The Ohio State University Psychological Examination was used in the Oberlin study. The average gain was 11 percentile points, from 73 as freshman to 84 as seniors. A critical ratio of 7.4 indicates a significant increase. The greatest gains were made by students in the lowest deciles. Many of the students in the lowest deciles dropped out of college before reaching the senior year but those who survived showed a significant group gain in intelligence. Hartson reported a negative correlation between initial scores and the amount of improvement. He concluded that improvement in intellectual abilities continues during the college years in those areas where stimuli are provided. Hartson interpreted his findings as an argument against the gradual unfolding of abilities by maturation independent of particular training.

THE PROBLEM

The present investigation was conducted to determine the amount of growth in intellectual abilities during the college years as revealed by the American Council on Education Psychological Examination. The data consist of freshman and senior scores for 74 students graduated from Lawrence College in June, 1938. The freshman tests were given

[3] Wright, Melvin B. "The Development of Mental Ability at the College-Adult Level," *Journal of Educational Psychology*, XXII (November, 1931), pp. 610-28.

[4] McConnell, *op. cit.*

[5] Hartson, *op. cit.*

early in September, 1934 and the senior tests were given at the end of April, 1938, making an interval of about 44 months between tests. On the re-test 37 students were given the 1933 Edition, 20 the 1934 Edition (the same test they had taken four years earlier), 16 the 1935 Edition, and one the 1937 Edition. The results were made comparable by interpolating all scores into 1934 percentiles. Freshman and senior percentiles were available for 36 men and 38 women.

THE RESULTS

The results of the re-test are presented in Table I. The mean percentile rank for the 74 seniors was 81.3 ± 1.42 in April, 1938 as compared to 72.7 ± 1.62 in September, 1934. The mean difference is 8.6 ± 2.15 percentile points, with a critical ratio of 4.0. It can be concluded with safety that these 74 students made a significant group gain in intellectual ability during their four years in college. The median gain (8.3 percentile points) is about equal to the mean gain, while the gains for Q_1 and Q_3 are slightly smaller than the median gain. It is interesting to note that both Q and S.D. are slightly smaller in the senior year than in the freshman year, although the decrease is not statistically significant.

The very fact that nearly half of the students studied were above Q_3 or the 75th percentile on the national norms as freshman means that their possibilities for gains in percentile rank were quite limited. It is well known that scores which fall around the middle deciles of the percentile distribution are more likely to fluctuate than are the scores at the extremes. When the 18 students who fell below 64.2 or Q_1, in the freshman year, are considered alone, their gain was 17.7 percentile points as compared to 8.6 for the total group. These 18 students had an average percentile rank of 42.6 as freshman and 60.3 as seniors. The amount and

TABLE I

CHANGES IN INTELLECTUAL ABILITY DURING THE COLLEGE YEARS FOR SEVENTY-FOUR LAWRENCE COLLEGE STUDENTS

Variable	Freshmen	Seniors	Differences
M	72.7±1.62	81.3±1.42	8.6
Q_1	64.2	69.2	5.0
Md	82.0	90.3	8.3
Q_3	91.5	95.8	4.3
Q	13.7	13.3	−0.4
S.D.	20.8	18.3	−2.5

direction of the changes made by the entire group of students during their college years can be seen more clearly when presented in tabular form. The changes found have been summarized in Table II.

The data in Table II show that the large majority of the students improved in intellectual ability during the college years. Some students seemed actually to have lost in intellectual competence, while others

TABLE II

THE AMOUNT AND DIRECTION OF THE CHANGE IN INTELLECTUAL STATUS

Gains and Losses	Men	Women	Total
40 and up	2	0	2
30 to 39	2	1	3
20 to 29	1	4	5
10 to 19	9	5	14
1 to 9	15	18	33
No Change	3	2	5
−1 to −9	3	5	8
−10 to −19	0	1	1
−20 and below	1	2	3
Total	36	38	74

made no significant change. Two students gained 40 percentile points or from 38 to 78 and from 44 to 84, respectively. These two men were so low in rank as freshmen that they were hardly considered good college risks, but as seniors one exceeded the mean of his class while the other was less than four points below the mean. Twenty-four students gained more than 10 percentile points, which seems to indicate a highly significant improvement for these students. Only four students lost as much as 10 percentile points, while eight lost from one to nine points.

Sex Differences.—It appears from the data in Table II that the largest gains are made more often by men, and the losses are more common among women at Lawrence College. This fact is presented more clearly when the gains and losses are expressed in percentage for each sex. The data in Table III reveal that the men made larger gains and had fewer losses. The reader's attention is called to the duplication in Table III. Four men gained 30 percentile points or more and 14 gained 10 percentile points or more. The 14 who gained 10 points or more include the four who gained 30 points or more. Likewise, the 29 men who made some gain include all men who gained whether their gain was one or 40 percentile points. Four out of five men (81 percent) made a gain while

three out of four (74 percent) of the women gained. About one out of three (32 percent) of the students gained 10 or more percentile points. Only five percent of the students lost enough to make their change in status significant. These findings seem to indicate that at least one-third of Lawrence College students who have the four-year college experience are likely to change significantly in their intellectual status. The men have the largest gains and the greater number of gains while the women suffered the greatest losses and accounted for a larger share of the decreases.

When the sex difference for the group as a whole was considered it was found that the women exceeded the men in percentile rank at both the freshman and senior levels. The men had a percentile rank of 69.6 ± 2.53 and the women a percentile rank of 78.7 ± 1.86 in the freshman year. The difference between the men and women was 9.2 ± 3.14 percentile points. A critical ratio of 2.94 indicates that this difference

TABLE III

THE PERCENTAGE OF EACH SEX MAKING VARIOUS AMOUNTS OF CHANGE

Changes	Men		Women		All Subjects	
	N	Percent	N	Percent	N	Percent
Gained 30 percentile points or more	4	11	1	3	5	7
Gained 10 percentile points or more	14	39	10	26	24	32
Gained 1 percentile point or more	29	81	28	74	57	77
Made no change	3	8	2	5	5	7
Lost 1 percentile point or more	4	11	8	21	12	16
Lost 10 percentile points or more	1	3	3	8	4	5

is not clearly significant statistically, but suggests strongly that Lawrence College men are, in their freshman year, somewhat inferior intellectually to the freshman women. In the senior year the mean percentile ranks were 79.4 ± 2.08 and 84.2 ± 2.00 for men and women respectively. The sex difference at the end of the college career was 4.8 ± 2.89 percentile points in favor of the women, with a critical ratio of 1.67. The men had reduced the sex difference in intellectual ability by approximately one-half in the four years. There is at least a possibility that the superior gain found among men is due to the later maturation of the male.

The men had an average gain of 9.8 ± 2.94 with a critical ratio of 3.33, while the women made an average gain of $5.4 \pm .86$ with a critical ratio of 6.26. The gain was much greater for men than for women but the significance of the gain made by women shows that there was a real

improvement during the college career. The greater variability among the scores for men makes a larger probable error and therefore a smaller critical ratio. There were only three women below the 50th percentile as freshman as compared to 10 men. In the senior year only one woman was below the 50th percentile as compared to seven men. The lowest percentile rank made by a senior woman was 20 as compared to 44 for the men.

Maturation.—The five students who gained more than 30 percentile points showed evidence of significant maturation during their college years. Four of these five students were men which lends further support to the possibility that the gain was due to a maturation factor. The only woman in this group informed the writer that she reached pubescence somewhat later than the average and considered herself a slowly developing person. One of the men who gained 40 percentile points reported that he grew 2 inches in height and gained about 25 pounds in weight during his college career. There is evidence from his college record to show that he matured significantly in social abilities as well as in intellectual and physical characteristics.

Relation of original status to change.—The correlation between the percentile ranks in the freshman and senior years was .82 ± .03. The size of this coefficient corresponds very closely to the result (.83) reported by McConnell[6] for Cornell College students. It must be concluded that there was little shifting in rank during the college years. Those students who were high when they entered college remained so and those who were low seemed to improve but failed to exceed their brighter colleagues. A negative correlation of —.49 ± .06 between the freshman percentile rank and the amount of change gives additional support to the finding that those students who were low in the freshman year gained most. It must be recalled that one-quarter of these students were in the upper decile (Q_3 equals 91.5) when they entered college and therefore could not have gained as much as the average for the group if each one of them had reached the 99th percentile in the senior year. The fact that the best students as a rule made some gain is shown by the increase in Q_3 from the freshman to the senior years. One-quarter of the seniors were in the top four percentiles (96, 97, 98, 99) while more than half of the class was in the upper decile (Md =90.3). The fact that such a large number of these students was so near the maximum percentile makes the gain of 8.6 percentile points for the group quite significant.

[6] McConnell, *op. cit.*

Those students who lost, ranged from the 30th to the 91st percentile freshmen. Nine of the 12 students who had losses were, however, in the middle 50 percent for Lawrence freshmen, 64.2 to 91.5. It seems probable that these students found themselves able to keep pace with their colleagues without exerting too much effort and fell into the habit of being satisfied with an average performance. Two of the greatest losses were made by students who were known at the time to be under great strain. It is likely of course that the minor losses, as well as the minor gains, were merely the result of the variability due to the test itself and therefore should be disregarded.

Field of the major and growth.—Experimental evidence has long since established the fact that intellectual growth in the secondary-school years is dependent upon factors other than the subject or combinations of subjects pursued by the pupils. It might be expected that the college years are not so dissimilar that the results would vary significantly. The existence, however, on certain course combinations and the apparent faith in certain subject fields has led the writer to tabulate the gains according to the field of the student's major. The size of the sample has necessitated combining certain majors. The writer has taken the liberty to combine the students into four major groups: Fine Arts, Language Arts, Science and Social Science. *Fine Arts* has been considered as a field in itself. *Language Arts* includes majors in English, French, German, Spanish, Latin, and Greek. *Science* includes majors in biology, chemistry, geology, mathematics, and physics. *Social Science* has been interpreted as including economics, government, history, psychology, philosophy, religion, and sociology. It is recognized that these combinations do not make entirely congenial families, yet the results may shed some light on the problem. It should be indicated that chemistry and biology account for most of the *Science* majors, English and French most of the *Language Arts* majors, and economics and history all but five of the *Social Science* majors. The gains are presented in Table IV.

TABLE IV

THE RELATION BETWEEN THE AMOUNT OF CHANGE AND THE FIELD OF THE MAJOR

Major Groups	N	Freshmen		Senior		Gains	
		M	Md	M	Md	M	Md
Fine Arts	8	88.8	89.0	96.6	95.0	7.8	6.0
Language Arts	25	74.6	79.0	82.9	87.0	8.3	8.0
Science	14	76.3	83.5	84.3	92.5	8.0	9.0
Social Science	27	69.2	72.0	74.8	82.0	5.6	10.0

It would be difficult on the basis of the evidence presented in Table IV to show that any of the large fields of study had superiority over any other large fields of study in contributing to intellectual development at the college level. It appears that colleges may soon be forced to accept the finding that any body of content is about as important as any other body of content for intellectual development. It seems clear to the writer that there are factors involved in presenting materials, stimulating learning, guiding the student, and challenging independent thinking which are of more importance than the particular subject matter on which the student does his thinking.

It is interesting to note that those students who elect to major in Fine Arts rank above students in any other field. The Fine Arts field at Lawrence College is not a series of courses pertaining to skills in music, art, and dramatics. It is rather a field of intellectual endeavor in which a student is challenged to think and study with the materials available in the Fine Arts. These students gained as much on the average as those majoring in the Language Arts and Sciences.

The Social Science majors ranked lowest and seem to have gained slightly less on the basis of the mean. The median gain in this group is, however, higher than that in any other field. The factor which produces this difference of 3.4 points between the mean and median for the Social Science group is that two students with the greatest losses are in this group. These losses reduce the mean significantly but do not affect the size of the median.

It would appear from the results of this study that colleges might well afford to spend more time considering the problems involved in giving students real intellectual stimulation and less time in counting hours and credits and arranging for supposedly desirable course combinations. It is highly probable that any professor who looks upon his courses as opportunities to stimulate growth and foster independent thinking rather that considering them as subjects to be learned will in the long run reap significant rewards in student appreciation for the stimulation given.

CONCLUSIONS

The results of this study of the intellectual growth of 74 Lawrence College students seem to justify the following conclusions:

1. The gain in intellectual ability of 8.6 ± 2.15 percentile points with a critical ratio of four indicates that there is a real improvement in intellectual ability during the college years.

2. Lawrence College men seem to be slightly below Lawrence College women in percentile rank as freshmen. The means are: 69.6 ± 2.53 for men and 78.8 ± 1.86 for women, giving a difference of 9.2 ± 3.14 with a critical ratio of 2.94.

3. The men gained more often than the women, had larger gains, and fewer losses.

4. The men were still slightly below the women in the senior year but the difference is not statistically significant. The means are: 79.4 ± 2.08 for the men and 84.2 ± 2.00 for the women, giving a difference of 4.8 ± 2.89 with a critical ratio of 1.67.

5. There appears to be no relationship between the amount of intellectual gain and the field in which the student majors.

6. It appears that at least one-third of the students in this study changed their intellectual status significantly. This fact presents a challenge to the college to make certain that each student is stimulated to the maximum of his potentialities.

7. The large gains made by a few students give a strong suggestion that at least some of the students who are admitted to college are a long way from intellectual maturity and therefore need careful guidance which will result in maximum development.

ISSUES IN THE CONSTRUCTION AND USE OF EDUCATIONAL MEASUREMENTS*

C. O. MATHEWS
Ohio Wesleyan University

Editor's note: The author presents six important issues in the construction and use of educational measurements with a caution that the case for measurement be not overstated.

I will not discuss critically all of the assumptions which have been enumerated by the previous speakers. The one thing which probably stands out in the minds of all of us is the confusion and variety of points of view which seem to exist. I shall organize my remarks around six outstanding issues which each of us interested in the construction and use of instruments for educational measurement must face.

1. *Shall the point of departure for the construction of valid instruments for educational measurement be a body of subject matter whose boundaries "are ascertainable with sufficient degree of exactness," or shall it be the outcomes of learning (psychological changes in individuals which have been carefully defined in terms of behavior)?* It will make a tremendous difference which of these points of view we adopt. If we begin with bodies of subject matter, we will sample by means of careful analysis the content of textbooks, courses of study, teacher examinations, and other such bodies of content. We will trust to this sampling and to a comparison of our test with other tests similarly constructed as the chief means of validation. Apparently the most central assumption underlying this whole procedure is that there is a common body of ideas to be learned.

On the other hand, if we make our point of departure the outcomes of learning, we will at once ask ourselves what changes in behavior we want to measure? How can these changes be described so that those people who have acquired the behavior can be distinguished from those who have not? Once we have clearly defined these psychological changes we will then search through any suitable available sources for means and materials for the construction of some device to secure evidence as to the extent of these changes. We certainly will not be interested, from this point of view, in limiting ourselves to knowledge acquired from certain bodies of subject matter, but will want also to secure evidences in regard

* A discussion before the American Educational Research Association, Cleveland, Ohio, February 28, 1939, of the contrasting points of view brought out by four speakers on one of the programs under the heading of "Fundamental Assumptions in Educational Measurements." The four speakers were: Dr. Frederick S. Breed, Dr. L. Thomas Hopkins, Dr. Ralph W. Tyler, and Dr. Ben D. Wood.

to changes in attitudes, interests, points of view, methods of work, and the like.

2. *Is behavior atomistic, mechanical, and additive in its nature, or rather is it unitary, organismic, and integrative?* In the last few years a good deal of evidence has been accumulating to show that behavior is not merely composed of a number of separate acts. The findings of Jennings, Childs, Cogwill, Wheeler, Lashley, and the gestaltists support this point of view by demonstrating that those descriptions of behavior which consist of merely piecing together the descriptions of numerous separate performances are inadequate. The integrative nature of behavior calls for a kind of measurement which will appraise behavior more nearly in terms of its totality. Perhaps we will have to compromise in this ambition to measure behavior as a whole, but we certainly have been guilty in the past of acting as if we had measured all of it when we have interpreted the results of our measures. Certainly the least we can do is to strive for more comprehensive measures and to try to be more scientific in our interpretations of the results of partial measures.

A technique which is rapidly developing, namely factor analysis, may result in certain conflicts with the integrative point of view. If we isolate through factor analysis more and more detailed independent factors and then through the measurement of each of these expect to get a measure of the whole, we may find ourselves in serious conflict with any adequate description of behavior. This issue reminds one of the sign hanging around the neck of the blind beggar standing in front of Grand Central Station in New York. This sign read:

Wars	2
Wounds	4
Battles	5
Wife	1
Children	4
Total	16

Then below this total were these words: "Help the Blind."

3. *Do local, state, and national testing programs tend to determine and fix the outcomes of instruction, or do these programs merely conform to and measure these outcomes?* When teachers have a choice in the matter they do not choose to use educational tests which they think measure outcomes widely different from their own. However, teachers frequently do not have the choice of such tests in their hands. Administrators who are interested in finding out whether their institutions rank

favorably with other institutions often make these choices. (Strange to say, this choice is often made without any very clear-cut conception on the part of either teachers or administrators of the objectives to be measured.)

This raises the issue as to whether or not there is, or necessarily should be, a significant "common core, varying in nature and magnitude from subject to subject, but of indispensable importance in all subjects in all schools which teach these subjects regardless of differences in marginal emphases." Our disagreements might be nearer an end if we would conceive this common core as a common core of objectives and not one of content. But even so, there is still the problem of measuring that part of a course or subject which is "common" to other schools and interpreting the results in the light of the school's individual goals.

There is considerable evidence that administrators and teachers make special effort to have their pupils and schools show up well on such tests. This means that the tests are controlling the outcomes and the emphasis on the time schedule. The desirability of having a school rank high in order to win a scholarship contest or to give a certain amount of prestige to that school and its faculty is a kind of extraneous motivation which may well divert education from its proper channels. If our tests were complete and valid measures of all desirable outcomes this would not be such an important issue.

4. *Is validity merely inherent in the instrument of measurement or is it a much broader concept than this?* Of course the content of the test, rating scale, questionnaire, or observation blank has much to do with the validity of the instrument. On the other hand, in a practical situation we are interested in the validity of the measure and not merely of the device utilized. The environmental conditions under which a measuring device is used also affects the validity of the measure. If a child is under intensive emotional disturbances, or if there are disturbing elements in the environment at the time the measurement is made, it may affect the validity of the outcome. As has been pointed out in recent research studies of child development, the stage of the development of the child may be one of these disturbing factors. The same score received by two children on an educational measure of some sort may warrant quite different interpretations depending upon the stage of their development. This raises the fundamental question as to just what it is that our educational measuring device actually measures. We cannot depend as much as we have in the past upon surface analyses but must understand more

fully the deeper meanings which affect the scores on such measuring devices.

As a part of this whole question of validity it may be well for us to consider whether or not we have placed too much faith in objectivity and reliability. It has sometimes seemed from reading literature in the field of educational measurements as though some authors placed these concepts on a par with validity. There really is no inherent value in objectivity and reliability as such. Our worship of objectivity has probably even limited us in our experimentation toward new forms of measuring devices. Thus, we have often been content with low validity and with a high reliability due to objectivity. It is necessary for us to understand the supreme importance of validity in this larger sense for unto that measuring device which has validity all things shall be added.

5. *Should not cross-section pictures of individuals and schools attained by survey and testing techniques give way to longitudinal pictures which will more validly reflect individual development?* We have seen a growing tendency both in the work of the Progressive Education Association and in that of the American Council on Education to place more emphasis upon continuous cumulative records of the individual. Any attempt to measure behavior which is recognized as unitary and integrative in its nature will demand such records so that the trends and nature of development may be determined from a more adequate measure of these characteristics than can possibly be furnished by one or a few unrelated measures.

6. *Have the outcomes of educational measurement (the changes which they have brought about in education) justified the claims made for them?* As a consequence of the development of measurement in education we have a wide-spread use of so-called homogeneous grouping, some rather deterministic types of guidance, an over-emphasis upon those aspects of learning which tests can measure, and a rather limiting sense of respect for objectivity.

As opposed to the above characteristics, which in many ways may be considered undesirable results of measurement, we have among a number of people in our profession an increased respect for the evaluation of outcomes, a desire for adequate data before the making of generalizations, and a recognition of the difficulties involved in trying to measure such a complex activity as human behavior. The very fact, however, that we in this room cannot agree upon a few fundamental generalizations or principles concerning the measurement of educational outcomes leads me to

believe that we cannot be too cautious about our use of these devices in the interest of better education. In spite of my rather negative criticisms I believe that the measurement movement is contributing worthwhile outcomes to education. We must be careful not to overstate our case. "Caution" is the watchword until we have arrived at a statement of principles supported by enough evidence so that we can agree upon the fundamental nature and value of educational measurement.

THE LAWS OF NATURE DETERMINE HOW TO MEASURE EDUCATIONAL PROGRESS

Glimpses of Mental Activity

GEORGE A. BROWN, C.E.
Bloomington, Illinois

ALTHOUGH children are not directly conscious of the fact, since it is of necessity a natural growth toward maturity, their minds are reaching out and asking questions in an effort to understand fellowship needs as they play and work together. What we may call educational progress, meaning by this name more than the action of the animal organism for its health needs, begins with awakened attention. The psychologists are able to detect responses which are beginnings of intelligent action in early infancy. These observed experiences may be given the double name of interest and effort. In this sense the name describes a direct cause of mental activities in the growing minds of children. Sensations are humanized in a way to cause the organization of ideas. The mind connects observed relations of persons and things and discovers a means of understanding. Interest and effort name the source of an energy which is characteristic of human life. The continued action of this energy in human life is educational progress.

The scientific name for this energy is human synergy. This name is borrowed from the science of biology where it names the energy which combines an interaction of the various organs of a plant or of an animal for the health of all, so that each coöperates to maintain the individual and the race. Human synergy combines the energies of individual minds in the interest and effort of each to advance civilized life in its social organizations.

In connection with this fundamental synergy (akin to Spearman's "g" factor common to all intelligence and here described as interest and effort) some abilities are gradually developed in the field of organized ideas. The fellowship experiences in this field are concerned not only with mutual understandings but also with responsive activities due to an urge for personal creative experiences which may or may not result in the immediate advance of a richer life for society.

The term "aptitude" may be applied to this organizing activity to express its more definite meaning as an energy than do the terms "capacity" or "general intelligence" and yet not so specific as the term "talent." As here used, aptitude names the mental energy that organizes

ideas through the process we call learning; that is the process of gaining an understanding from related experiences and ideas. This energy of aptitude holds an attentive interest to certain related fields of experience in connection with ideas and gets a clear perception of the facts in their relation to organizing principles. This direct study finally provides the means of knowledge and skill acquired in response to the energy of aptitude, and the educational process organizes ideas for use in the various fields of human interests.

The question of a measuring instrument adapted to so complex a development of natural energies is today being considered from various angles of approach. The measurement instrument question, however, is related only to the mathematical device best suited to give as definite a measure as possible. In the physical sciences, such devices are related to simplifying processes of computation, and, where needed, to indicate combined relations by certain signs; such as the "–" for negative number and the vector sign "i." The decimal and integral calculus provides for the study of strains and stresses in structures distributed between positive and negative action from a common center. If any device such as this is needed for measuring educational progress, its purpose will be, if possible, to measure the results of the action of the law of nature due to growth toward maturity, and also the action toward fruition in mature fellowship. Vector analysis provides for combining in the same symbol measures of known direction and of quantity. Does the law of aptitude-development involve this combined action? This question is asked simply to center the discussion on the concept of natural law in the process of its development of an on-coming generation. Each individual has the urge of synergy, in the form of interest and effort action on brain equipment, by which sensations take on the form of more or less permanent ideas. Also, in connection with and as a part of this ability, further social growth includes the organization of ideas with the equipment of fellowship relations, to the end that aptitude leads to mutual understandings and to the need to learn facts and conditions concerned with coöperative activities.

In the processes of learning there is what may be described as a back and forth movement of ideas and ends by which understanding is broadened and knowledge and judgment perfected. This action of aptitude belongs to the educational process as it uses the learnings which human relations, in a free union of common efforts, have developed. In contrast, the strong action of a particular talent in one field of expression may

have little interest and effort for the wider development of the underlying aptitude for learning in relation to more common needs of society. Talent emphasizes its own expression as, perhaps, a gift to society. As an enrichment to life, this may encourage more interest and effort for its expression in many communities.

The school situation has recognized this need to nurture the growth of power in close relation to growing maturity for the uses of aptitude energy. There is a decided contrast between the nurture of growth of a power energy and that of discovering new uses of a physical law, the action of which is already established. What, then, are the essential uses of measurement as a guide to educational progress? Evidently measurement, in relation to the nurture of growing abilities of a nature energy, is closely connected with measurement in relation to progress in getting clear ideas of facts in their relation to organizing principles. It should be possible to find a common unit of measure for the two relations. Moreover, the underlying energy of synergy is common to both situations. The most striking difference in the measuring conditions is, that the nurture of individual growth of nature-given energy is concerned with a natural condition of endowment which determines limitations for one and extended opportunities for another. Knowledge of this condition is of the greatest importance for those concerned with child development. The value of measured abilities is so clearly important to the school teacher, as she guides children with various degrees of age growth, that this measurement should be available to guide her work with individuals in all classes.

Measurement in the field of educational progress in learning has relations to most of the school conditions and to all who are directly responsible for them. These relations center in the provisions for group coöperation in all learning, and in provisions for skill activities. Nature's law of aptitude growth is that of an energy which works for an understanding obtained by organizing ideas and experiences toward agreement within themselves, and with accepted facts of common knowledge. It includes comprehending and judging. Aptitude perceives the significance of agreement or disagreement between ideas and facts in determining means to ends. Reflective thinking is required and should, of course, be nurtured as part of the educational process; and this should be done in a way to advance from simple and easily seen objective relations toward the subjective situations of the perception of contrasting ideas. School conditions should be adjusted from year to year to the need for this nurture.

EDITORIALS

SIGNIFICANT TRENDS IN RESEARCH IN DIAGNOSIS IN ARITHMETIC

THERE is an abundance of empirical evidence which shows that remedial measures applied to the removal of factors interfering with learning in progress will result in success if these measures are applied with discrimination. This has been well demonstrated in such areas as reading, arithmetic, and language. It must be admitted, however, that there is little direct evidence as to the particular types of remedial measures that are useful in dealing with specific handicaps nor do we know with any degree of assurance what factors or combinations of factors contribute to the development of particular types of unfavorable growth of the learner.

It is highly significant that much of present day research in learning is being directed to the identification and prediction of handicaps at early stages of learning so that ultimately steps may be taken to prevent them through the adjustment of the educational program to the needs of individual learners. These investigations may be grouped under four heads namely (1) studies in the field of readiness, (2) studies of the characteristics of good and poor achievers, (3) studies dealing with the more effective organization and gradation of the contents of the curriculum, and (4) studies dealing with the ways in which learning takes place at various stages of the learning process.

Studies in the field of readiness have largely been limited to the prediction of success in reading and arithmetic in the primary grades. There is clear evidence that the tendency in readiness testing is to shift from mere prediction of success to the identification of factors that may interfere with learning. There is no reason why these two purposes of prediction and diagnosis should not be served by a well constructed series of readiness tests. It has also been pointed out by several writers that readiness testing need not be confined to the initial stages of learning but should be extended to higher levels. The possibilities in this direction have been demonstrated by a recently published series of readiness tests in arithmetic processes for grades 2 through 7. This approach should be extended to other aspects of arithmetic, for example, social arithmetic,

the study of such topics as area, volume, and other formal phases of the subject. Lack of readiness of pupils has been amply demonstrated by these tests and is undoubtedly a cause of failure to progress satisfactorily that can be remedied by adjusted instruction.

In order to aid in the location of factors that contribute to unfavorable development of learners, a number of important investigations, notably those of Gates and his associates, have been conducted to determine differences in the characteristics of good and poor achievers in the various areas of the curriculum, primarily in the field of reading and spelling. A valuable body of information has been gathered which is highly useful in organizing educational programs. The implications of these researches are well stated in the statement of Gates that the ideal procedure is "first and primarily individual prescription for individual needs." This type of research leading to the identification of causes of poor achievement should be extended to all areas of learning. It is particularly adapted to the field of arithmetic.

The obvious importance of gearing the curriculum to the developmental process so as to facilitate learning has led to numerous studies of the most effective ways of organizing the curriculum. The pioneer studies of the Committee of Seven in arithmetic are being extended and checked by other investigators. There is ample evidence of a wide range of difficulty of the various elements of number processes, a fact that raises important questions as to how best to organize the curriculum for successful learning. When should a topic be first introduced? What experiences should be provided to get the child ready for the new topic? Over how long a period should the introduction of various phases of a topic be spread so as to facilitate and insure learning? How can the curriculum be organized so as to provide most adequately for differences in the rates at which children learn? These and numerous related questions are unanswered and their solution offers a rich field for research.

Closely related to the problems of organization of the curriculum are issues related to methods and materials of instruction. Effective classroom procedures should clearly recognize information that is slowly being accumulated about how learning takes place. The studies by Brownell and others show that learning is a gradual process that passes through various stages of immaturity until mature procedures result. This finding is especially significant in the field of arithmetic since it has been the common practice of teachers to expect pupils to acquire mature methods of reacting at the start. The fallacy of this assumption has been demon-

started by evidence of the value of procedures which stressed meaning during the learning process leading ultimately to mastery of desirable methods of work by pupils. It is a fact worth noting that in these studies the techniques that were devised to study errors made by pupils have been used to make longitudinal studies of the learning process under way. This is an interesting illustration of the intelligent adaptation of a procedure developed for one purpose to another purpose for which it was ideal. More studies should be made of the nature of the learning process, how learning takes place, and the influence of various instructional procedures that condition learning so that teachers can more intelligently direct the learning activities of their pupils.

LEO J. BRUECKNER

MORE ABOUT REVIEWS OR ABSTRACTS OF EDUCATIONAL RESEARCH

IN the November issue of the *Journal of Educational Research* there appeared as written by the undersigned an editorial, the heart of which may be summed up in two sentences:

> Those preparing reviews of educational research should not report the investigations which are obviously so faulty in procedure as to make the conclusions unreliable. To report to teachers and school administrators such studies without sufficient detailed description of the sources of the data and the techniques of treating them so, is unethical. (Page 302.)

In the December issue of the Journal, Dr. Samuel W. Fernberger, an eminent psychologist takes the position that to fail to report every study made in a field is unethical.

There is considerable merit to Dr. Fernberger's statements. I am afraid however that he did not visualize clearly the situation I had in mind. Dr. Fernberger was apparently thinking of abstracts of studies which were in reality condensations of individual studies. What I had in mind more particularly was reviews of research studies in which the individual researches, particularly its technical procedures, were not described in detail. It is unfortunate that in the title of my editorial I used the word 'abstract.''

Dr. Fernberger is apparently thinking of abstracts prepared for other scientists in the field concerned. I am thinking, primarily, of reviews intended for field workers who would apply the results of research. Dr.

Fernberger says, "But one expects the reader of an abstracting service to read in the original those articles that interest him or that may be valuable in his field of work." One may know without the slightest shadow of a doubt, that not only will the very large majority of readers of reviews of educational research, teachers and school administrators never see the original articles, but the majority of them could not understand nor evaluate the technical procedures and terminology employed. To pass on as truth to these people the results of investigations which by accepted canons of valid procedure are clearly faulty and unreliable, and thus to mislead them, is not only unfortunate, even unethical, but little short of quackery and charlatanism.

If an abstract or a review of educational research is neither critical nor selective or if done by one who is not competent to judge of the reliability of the procedures of investigation, the reader should be so warned and the wording of the review or abstract should be such as to avoid carefully any suggestion that the conclusions of the investigation may be trusted. In other words it should be little more than a rather full annotated bibliography of research in the given field.

Let us take an actual case, in which I prefer not to use names. Mr. A, a graduate student, in his M.A. thesis reports that there is no difference in the relative efficacy of two methods of teaching history of the United States. A reading of his thesis reveals that one of the methods was a comparatively novel one with which the teacher had little previous experience. Two classes of approximately 20 pupils each were employed in the experiment. Of these, 24 pupils in each class were mutually paired for the purpose of holding constant certain variables particularly the learning abilities of the two groups of subjects. The pupils are paired on the basis of their I.Q. on a certain intelligence test. Their progress was measured by an objective test requiring 90 minutes. A reviewer reports that Mr. A found no difference between the relative efficacy of the two methods as actually investigated under experimental conditions with two eleventh grade classes. In the review he does not note that (1) since I.Q. and not M.A. was held constant that: (1) one class may have been more capable of learning than the other; (2) one method was relatively handicapped by the lack of experience of the teacher with it; (3) the investigation does not even describe the method well enough to enable us to know what the experimental variable actually was; (4) since no measure was had of the abilities possessed by the pupils at the beginning of the experiment period, the scores on the test at the end of the experiment period probably do not

represent the amount of growth; (5) the rest were tests merely of subject-matter information of the detailed type and afforded no means of measuring some of the important objectives of teaching history. In sum the reviewer after assuming a role of technical competency in which the reader is encouraged to have faith, passes on to untrained and unsuspecting teachers conclusions, which they will accept for facts, which are for all their scientific window dressing probably no better than a good guess.

For anyone to hold that this sort of disservice should be continued, that it serves any good end, places a severe strain upon logical processes.

Let us take another example, Dr. B in a study of what factors are associated with the relative effectiveness of teachers focused upon two classifications of teachers—those at the upper end of the distribution—the relatively successful ones and those at the lower end, the relatively unsuccessful ones. For each teacher in these groups he has data concerning the number of years of Latin studied in high school. He takes as a measure of the relationship between success and number of years of Latin, biserial r, a procedure clearly inapplicable and certain to yield a coefficient of relationship much higher than that obtained by valid procedures. A few years later a man of international reputation and enormous following among teachers in this country, most probably mislead by a review of the study, for this great man does not ordinarily concern himself with more complicated statistical studies, seized upon this unwarranted conclusion of the study to support his point of view that the study of Latin more than the study of other subjects develops qualities making for teaching success.

Cases of this sort could be cited by the score. For anyone to maintain that to report uncritically such studies is anything but a disservice is difficult for me to imagine. Dr. Fernberger need have little worry that by reason of the failure to report these amateurish researches, the loss of valuable discoveries such as Mendel's work in genetics may be experienced. What is more unlikely is that in this day of uncritical attitudes towards this huge amount of research of varying degrees of competence, is that studies of real value will be ignored rather by reason of the fact that everyone is busy doing research that few have time to read it and those few read but a small part of the huge and increasing amount of it and uncritically with but poor understanding.

Dr. Fernberger says "the failure of a review to abstract an unpublished paper, if he knows it exists, is unethical." His discussion would seem to suggest that I condone or approve the failure to report such

papers. I did not discuss the issue. I believe with Dr. Fernberger, that a reviewer of the research on a certain point or problem should examine all the studies available whether published or unpublished, whether easy to get at or not. Under the conditions that prevail, we are not doing that in education. Reviewers do their work, for the very great part, on their own time as a labor of love and on their own expenses. In the successive issues of the *Review of Educational Research*, it is the exception if an unpublished thesis is reported. Yet in the bulletins of the United States Office of Education, there are listed annually more than two thousand M.A. and Ph.D. theses in education carried on in this country alone. As a result of these omissions as well as the uncritical and superficial reviewing of educational research, teachers and others who wish to know what research has to offer them tend to rely too trustingly or perhaps too suspiciously upon "reviews of research."

It is indeed unfortunate that in these days of granting hundreds of thousands of dollars by the foundations to projects which even if successful are of so questionable value to education, that no funds are available for the payment of salaries of competent technical experts to review educational research and to report its most probable sound conclusions to the hundreds of thousands of workers in education. Shortly there will appear an encyclopedia of educational research. No funds were available to employ an adequate staff of critical and competent abstractors and reviewers. Topics were distributed to scores of men and women who on spare time and with whatever technical competence in statistical and research methods they possessed, prepared reports in their fields. Serviceable as this encyclopedia will be, how much more reliable and useful this work might have been if funds had been available to finance exhaustive and critical reviews.

What is needed in education by teachers, parents, and administrators, is a service that is now not available. They do not need reports of all attempts at research giving us, impartially, conclusions of loose, amateurish gestures at research, but reviews and syntheses by interpreters who are competent and who have available the time to search exhaustively for all research done on each problem, to scrutinize each investigation critically, to digest and evaluate it and to tell us all as far as can be told, what we really know about the thousands of unsettled problems that hold us back in education. May the time come when that service is available.

HARL R. DOUGLASS

REVIEWS

Address all communications relative to book reviews to Editor Kai Jensen, University of Wisconsin, Madison, Wisconsin.

HARRISON, M. LUCILE. *Reading Readiness.* Revised Edition (Boston: Houghton Mifflin Co., 1939), pp. vii + 255. $1.40.

The first edition of this book, which appeared in 1936, was concerned with problems of reading readiness in the primary grades. This material, with little change, appears in the revised text as Part I under the title of "readiness for the initial period of reading instruction." Items considered include (1) factors which influence reading readiness, (2) instruction to foster reading readiness, (3) the testing program, and (4) suggestions for placement and remedial work. Mental age is correctly considered an important determinant of reading readiness. It is advised that all children be provided with a rich and varied experience as a preparation for reading. Unfortunately, as in the first edition, Harrison tends to accept uncritically the claims made by authors of tests.

Part II of the text concerns itself with reading readiness at educational levels beyond the preparatory level. Although in general all the higher levels are supposed to be covered, most of the discussion is confined to grades below the high school. The chapter on the nature of reading is exceptionally well done. Undue emphasis upon the mechanics of reading and other narrow viewpoints, such as the development of rhythm reading, are severely criticised. In defining reading, the author makes use of desirable elements from many points of view. Due consideration is given the important part played by recognition of symbols which serve as stimuli for recalled and constructed meanings and the manipulation of these "meanings in thought processes according to the purposes of the reader . . ."

The requisites for reading the material at levels beyond the preparatory period and the instructional program for readiness to carry out the thinking activities of reading are outlined. Provision is made for the development in pupils of the necessary concepts for constructing meanings in terms of the specific material to be read. This may be done through direct experience or through vicarious or indirect experience. The former is preferred. Other parts of the program to develop readiness include making provision for vocabulary development, understanding of sentences, and reading according to purpose.

It is suggested that every subject-matter teacher should also be a teacher of reading since they are the ones best fitted to guide the student in the thinking side of reading such material.

Extensive bibliographies on research reports and on materials for use in the reading readiness program are given in the appendix. Also problems and exercises are given for material covered in each chapter.

This excellent book will help give due prominence to the problem of reading readiness at the higher levels as well as in the primary grades. Because a program of reading readiness at all educational levels is basically sound, this movement will without doubt become more widespread in the near future.

MILES A. TINKER

University of Minnesota

RESEARCH ABSTRACTS and BIBLIOGRAPHIES

Address all communications relative to research abstracts and bibliographies to A. S. Barr, University of Wisconsin, Madison, Wisconsin.

BULLETINS

ANDERSON, MIRIAM SYLVIA. "The History of Secondary Education in Waldo And Piscataquis Counties in Maine," *The Maine Bulletin*, XLII, No. 4 (Orono, Maine: University Press, 1939), 111 pp.

The purpose of this study is to trace the development of two Maine counties to show influence of the historical, social, and political conditions of the various periods upon the educational attitudes and accomplishments of the people of these two areas.

BETTS, EMMETT A. *Lateral Imbalance*, Data on Visual Sensation and Perception Tests, Part I (Meadville, Pennsylvania: Keystone View Company, 1939), 27 pp.

The author presents previously unpublished data on the reliability and validity of the readings taken on the lateral imbalance slide of the Visual Sensation and Perception Tests of the Betts Ready to Read Battery.

Educational Research Service. *Education in Lay Magazines, December 1, 1939*, Circular No. 11 (Washington, D. C.: American Association of School Administrators and Research Division of the National Education Association, December, 1939), 16 pp.

This issue of the Educational Research Service contains digests of articles taken from lay magazines and dealing with: teacher education a vital problem, a program for American education, the child and the school, the future of the endowed university, our schools: a symposium, and miscellany.

LUNN, J. E., ET AL. *Pupil Personnel Study of Pupils in Minnesota Public Schools*, Part I, Sections 11-15 (St. Paul: Minnesota State Department of Education, 1939), 171 pp. Mimeographed.

Presenting the results of studies of pupils in Minnesota schools, these five bulletins include: comparison of dull children and bright children as problem cases, the relation of certain personality traits to other educational factors, comparison of rural and urban pupils in the ninth grade, relation of environmental conditions to educational results, and relation of comprehension in reading to other factors in education.

REINHARDT, EMMA, BEU, FRANK A., AND HAEFNER, RALPH. "Changes in the Student Body of the Eastern Illinois State Teachers College During the Ten-Year Period 1925-26 to 1935-36," *Eastern Illinois State Teachers College Bulletin*, No. 147 (November 15, 1939), 52 pp.

The authors present the results of a study of the changes in the social, intellectual, and scholarship characteristics of students at Eastern Illinois State Teachers College, during the ten year period, 1925-26 to 1935-36.

Research Division. *Educational Interpretation* (Washington, D. C.: Research Division, National Education Associa-

tion, November, 1939), 17 pp. Mimeographed.

This bulletin contains a list of 204 references to help the educator interpret to the public the purposes and achievements of the public schools.

Research Division. *1939 State Legislative Progress of Teacher Retirement (to November 1, 1939)* (Washington, D. C.: Research Division, National Education Association, December, 1939), 7 pp. Mimeographed.

The Research Division presents an analysis of teacher retirement laws passed in state legislatures during 1939.

Research Division. *Status of Teachers' Salary Schedules for 1939-40 in Cities over 100,000 in Population* (Washington, D. C.: Research Division, National Education Association, October, 1939), 26 pp. Mimeographed.

This bulletin contains data from an inquiry sent in September 1939 to superintendents of schools in cities over 100,000 in population.

Research Division. *Bibliography on Youth Problems* (Washington, D. C.: Research Division, National Education Association, November, 1939), 38 pp. Mimeographed.

This bibliography provides 250 references on: the status of youth, earning a living, personal development, keeping healthy, going to school, belonging to society, religious growth, foreign youth, and other bibliographies.

SMITH, HENRY LESTER, AND EATON, MERRILL THOMAS. "An Experimental Study of Standards for the Selection of Candidates for Teacher-Training Institutions," *Bulletin of the School of Education Indiana University*, XVI, No. 1 (December, 1939), 72 pp.

After reviewing previous investigations the authors present the results of the study of the pre-training selection of teachers at Indiana University from which they conclude that the quality of those entering teaching can be improved through selection.

SWEETMAN, EDWARD. *History of the Melbourne Teachers' College and Its Predecessors*, Australian Council for Educational Research, Series No. 57 (Melbourne: Melbourne University Press, 1939), 144 pp.

The author presents a historical survey of the personalities associated with and the policies of the Melbourne Teachers' College, Australia.

TYRON, CAROLINE MCCANN. *Evaluation of Adolescent Personality by Adolescents*, Monographs of the Society for Research in Child Development, Volume IV, Serial No. 23, Number 4 (Washington, D. C.: National Research Council, Society for Research in Child Development, 1939), 83 pp.

The author presents a detailed analysis of data from the California Growth Study relative to the evaluation of adolescent personality by adolescents.

WOOD, BEN D. *1939 Fall Testing Program in Independent Schools and Supplementary Studies*, Educational Records Bulletin No. 29 (New York: Educational Records Bureau, January, 1940), 50 pp.

This bulletin contains a summary of the test results of the Bureau's fall testing program, and articles on: the reliability of the American Council on Education Psychological Examination, 1939 Edition; a study of the new edition of the Iowa Silent Reading Test for high schools and colleges; and some data on the revised Strong Vocational Interest Blank.

EDUCATIONAL ADMINISTRATION AND SUPERVISION

GOETSCH, HELEN B. "Relation of Parental Income to College Opportunity,"

The School Review, XLVIII (January, 1940), pp. 26-33.

The author presents data to show that the attendance of bright boys and girls at college depends upon the economic status of the parents.

CURRICULA AND OBJECTIVES

BRINK, WILLIAM G. "High-School Pupils' Interests in Magazines and Newspapers," *The School Review*, XLVIII (January, 1940), pp. 40-48.

This is the second of two articles reporting the results of a recent survey of the reading interests and activities of 1532 high school pupils in Chicago.

PSYCHOLOGY OF LEARNING AND TEACHING

CARPENTER, HARRY A. "Do Science Radio Broadcasts Pay?" *Science Education*, XXIII (November, 1939), pp. 299-304.

The author presents the results of a questionnaire study of radio in the Rochester public schools in which he finds that teachers are favorable to radio and have as a result of its use made important changes in their teaching.

OLSON, MYRON S., AND KAMBLY, PAUL E. "A Comparison of Three Types of Teacher Activity in Directing the Study of General Science," *Science Education*, XXIII (November, 1939), pp. 304-308.

The authors present the results of an experimental study of three methods of teaching science, namely the "individual," "group," and "help requested" methods from which they conclude that the individual is most effective.

WINTERS, ELWOOD J. "The Determination of the Meanings Which Students of Science at Different Grade Levels Associated with Selected Scientific Concepts," *Science Education*, XXIII (November, 1939), pp. 331-335.

The author presents data from which he concludes that either too much time is given to the memorization of facts or it is more difficult to instill scientific attitudes.

MEASUREMENTS, STATISTICS, AND SCIENTIFIC TECHNIQUES

GRANT, ALBERT. "Results of Nelson Silent Reading Test in Grade IX," *The School Review*, XLVIII (January, 1940), pp. 34-39.

This article presents an analysis of the results of the Nelson Silent Reading Test for more than three thousand ninth grade pupils in the Cincinnati public schools.

BIBLIOGRAPHIES

GRAY, WILLIAM S. "Selected References on Teacher Education," *The Elementary School Journal*, XL (December, 1939), pp. 297-306.

A selected list of references for the period July 1, 1938 to June 30, 1939.

RESEARCH NEWS and COMMUNICATIONS

Address all research news and communications to Carter V. Good, Teachers College, University of Cincinnati, Cincinnati, Ohio.

Plans for Completion of the Dictionary of Education.—During the Christmas holidays, at the Chicago meeting of the National Council of Phi Delta Kappa, plans were approved for completion of the Dictionary of Education. The sequence of steps for the next two years is approximately as follows:

1. Completion of the preliminary work of definition by approximately 150 coördinators or specialists and their assistants, working in as many subdivisions of education, by December, 1940.

2. Submission of definitions for criticism to other experts in the appropriate fields early in 1941, followed by any necessary revisions, to be made by the original authors of the definitions.

3. Editing of copy for release to the publisher by December, 1941.

4. Proof reading early in 1942.

The National Council also authorized the National Executive Committee to sign a contract with McGraw-Hill Book Company for publication and sale of the Dictionary, which will be made available at a special discount to members of Phi Delta Kappa. Royalties will accrue to the fraternity in return for the support given to the project.

Origin and Sponsorship of the Project.—Phi Delta Kappa voted at the Cincinnati meeting of the National Council (December, 1937) to sponsor preparation of a comprehensive educational dictionary. This action had been preceded by a period of preliminary work to which a number of individuals and organizations contributed. The Dictionary of Education is under the supervision of a committee (appointed by the Executive Committee of Phi Delta Kappa) whose membership consists of: Carter Alexander, Columbia University; W. W. Charters, Ohio State University; Paul M. Cook, ex-officio, executive secretary, Phi Delta Kappa; H. D. Rinsland, University of Oklahoma; W. Ray Smittle, Wayne University; and Carter V. Good, chairman, University of Cincinnati, who is in charge of the editorial office at Cincinnati.

Purpose and Contribution.—The primary purpose is to prepare a comprehensive dictionary of technical terms in education which will do for educational workers and teachers what already has been accomplished by technical dictionaries for practitioners in such special fields as medicine, law, engineering, and psychology. The Dictionary will promote greater clarity and consistency of usage in educational literature and in classroom instruction. Since the plan is one of sponsorship by Phi Delta Kappa, through a supervisory committee, and of coöperative participation in the definition of terms by coördinators or key men in various chapters, who are enlisting the aid of their associates, advanced classes, and seminars, the stimulation given to local chapters and to individuals through coöperative effort should be as valuable to Phi Delta Kappa as is the Dictionary itself to the field of eduaction.

The Master List of Words and the Handbook.—The first step was the preparation of a master list of words and a handbook of procedure, under the supervision of the Committee on Dictionary of Education. The list of words has been mimeographed in the form of a 122-page booklet, including some 19,000 entries. The 22-page handbook of information also has been issued in mimeographed form. It deals with the Dictionary project in terms of origin and sponsorship, purpose and contribution, scope and delimitation, a description of the master list of words, subdivisions of education included, illustrative definitions, bibliography on educational terminology, and matters of form. The list of terms and the handbook have been placed in the hands of coördinators or specialists representing some 150 subdivisions of education, and also have been distributed to the various officers and chapters of Phi Delta Kappa.

At a later date the complete list of 150 or more specialists will be made available through the columns of the *Phi Delta Kappan*. A few of the fields are not yet finally assigned to coördinators, due to inability to accept on the part of some specialists and difficulty in locating experts for certain other areas.

Coöperative Study in General Education.—The Coöperative Study in General Education, operating under the sponsorship of the American Council on Education, involves the coöperation of twenty-two colleges. The general aims are to assist faculties in redefining aims; to assist faculties in evaluating their program of general education; to develop within the coöperating institutions faculty leaders capable of stimulating and conducting internal programs of evaluation and curriculum research; to demonstrate the value of coöperative efforts among educational institutions; and to provide a means for the effective dissemination of the results of educational experimentation.

The twenty-two colleges work coöperatively. The Coöperative Study is not a survey, nor is it merely a study of specified aspects of general education. The Study concerns itself with problems that faculty members in coöperating institutions identify and define, not with ready-made lists of problems. The activities of the central staff grow out of the expressed needs of faculty investigators in the colleges. In a real sense members of the staff aim to serve as assistants to faculty investigators. Faculty requests determine the direction of central staff activities.

One major purpose is to explore the possibilities of various coöperative procedures. Colleges have common problems. The coöperative study of such problems should serve the mutual interests of all concerned. One means of coöperative attack is the establishment of inter-college committees to work on various aspects of common problems. Another is a provision for summer workshop opportunities in which faculty members spend five weeks together in the concentrated study of individual problems. Furthermore, members of the central staff visit the coöperating colleges upon the request of the institution. In some instances, faculty leaders in the coöperating institutions spend six months in the office of the central staff working as fellows. This experience provides training for the professor who will later assume leadership in the program of faculty studies of general education in his own institution.

The foregoing emphasis upon the coöperative study of common problems does not indicate any lack of concern with problems which are peculiar to an institution. If a coöperating college is interested in a problem which concerns that college

alone and if the college regards this problem as important, the resources of the central staff are available to that college as it studies its unique problem of general education.

At a two-day conference of liaison officers of the coöperating colleges, the group unanimously agreed that topics for special emphasis for 1939-40 should be the following: (a) study of student needs; (b) development of a comprehensive program of evaluation; (c) study of student personnel administration; (d) development of administrative procedures within the institutions to facilitate the Coöperative Study; (e) coördination of Coöperative Study activities with related activities of other national groups; and (f) development of clearing-house service.

Ralph W. Tyler directs the Study. Other members of the staff include Ralph W. Ogan as Associate Director, research associates, and research assistants. The research associates in science, humanities, and social science have their major graduate education and experience within the subject-matter fields in which they work. None is an educationist in the strict sense of the term. All of these men became interested in research in problems of teaching through their college teaching experiences. Each believes that there is much necessary research to be done in this field, just as in the field of the subject-matter discipline itself.

The office of the central staff is located at the University of Chicago, 6010 Dorchester Avenue, Chicago.

Evaluation of School Broadcasts.—During the first two years of the Evaluation of School Broadcasts investigations, the staff has constructed tests and other instruments of appraisal to evaluate changes in attitudes, interests, ways of thinking, information, and selected skills, as these have been influenced by listening to and utilizing school broadcasts in music, science, social studies, and literature. During the past year, the staff has designed and administered controlled experiments to evaluate selected outcomes for local, regional, and nation-wide broadcast series.

Studies of local broadcasts included the evaluation of a literature series in Chicago, a music series and a systhesis-of-school-learning series in Detroit, a music series in Rochester, New York, a science recording series in Cicero, Illinois, and a news broadcast series in Rochester and Detroit.

Studies of regional broadcasts included a music series in California and other coast states, the Standard School Broadcast, WLS "Music in America," Chicago, and a conservation series of the Ohio School of the Air, over Station WOSU, Ohio State University.

Studies of nation-wide broadcasts included selected program series of the Columbia Broadcasting System and the National Broadcasting Company. From the Columbia Broadcasting System's American School of the Air, the music series, science series, social studies series, and literature series have been investigated. For the National Broadcasting Company the Music Appreciation Hour (Damrosch), Series A, the lessons in the playing of orchestral instruments (Maddy), and the social studies series entitled "Ideas That Came True" were investigated.

An extension of the Evaluation of School Broadcasts studies for a three-year period has led to the development of new plans and projects. As these plans are developing, three major types of studies will be carried on. They include a Community Study, Wide-Sampling Studies, and Special Studies of selected local school broadcasts.

The Community Study has been designed to permit very intensive and extensive studies of the effects of in-school broadcasts upon school administration, supervision, teaching practices and pupil attitudes, information, ways of thinking, and emotional adaptability. Such broadcasts include Columbia Broadcasting System's "Frontiers of Democracy," "Folk Music of America," and "Tales From Far and Near." For similar purposes selected out-of-school broadcasts such as "America's Town Meeting of the Air," children's serial programs, and "Gang Buster" types of broadcasts will be studied. Such effects as improvement in critical thinking, changes in social, political, and economic attitudes, and the emotional adaptability will be related to sociological and psychological factors in the pupils' backgrounds.

The Wide-Sampling Studies have been designed to permit extensive rather than intensive studies of the effects of the same broadcasts on children of school age. In order to judge the validity of generalizations which may grow out of an intensive study of a single community, similar studies will be carried on in the schools of Los Angeles County, California; Houston, Texas; Denver, Colorado; Detroit, Michigan; Rochester, New York; Toledo, Ohio; Akron, Ohio; and Trumbull, Connecticut.

The Special Studies of local broadcasts have been designed to permit evaluation of the effectiveness of programs produced and utilized under the direction of local boards of education in Chicago, Illinois; Detroit, Michigan; Cleveland, Ohio; and Rochester, New York. The Evaluation of School Broadcasts staff, located at Ohio State University, will coöperate by gathering evidence which will permit local school officers to appraise and to improve their broadcasts. I. Keith Tyler is director and J. Wayne Wrightstone, who prepared this progress report, is associate director.

Teacher Examination Service. — In order to make available to school superintendents the best resources and devices now known for selecting teachers and to promote research for the improvement of teacher selection methods and devices, a teacher examination service has been organized by the American Council on Education with the aid of a grant from the Carnegie Corporation through the Carnegie Foundation for the Advancement of Teaching. The American Council on Education has undertaken the project after consultation with a group of superintendents, some of whom have been using certain tests of the Coöperative Test Service for several years as a part of their total selective procedure. These superintendents believe that there is a crucial need for an examination service which will avoid the costly duplications and uncertainties of locally assembled examinations for teacher candidates; which will give reliable and valid measures; and which will be comparable from year to year and from one school system to another. They point out that the need for comparable measures of the intellectual status of candidates for admission to the teaching profession has been clarified and accentuated by the revelations of the Carnegie Foundation's Pennsylvania Study, the annual reports of the Teachers College Personnel Association, and the American Council's testing programs. While this examination service is designed primarily to aid superintendents in their efforts to improve the selection of new teachers, it is obvious that the opportunity to "register" talents on a national scale, in terms of widely comparable units, will have very great advantages for the candidates and institutions preparing teachers.

The American Council has placed the teacher examination project in charge of

a National Committee on Teacher Examinations, of which Supt. A. J. Stoddard of Philadelphia is chairman. The committee plans to develop a three-fold service. First, the committee will make available to the cities coöperating in the experiment a battery of carefully constructed tests. The task of constructing, administering, and scoring these examinations, and of reporting the results, has been assigned to the Coöperative Test Service, of which Ben D. Wood is director. Second, the committee will supply as much information as possible to the coöperating school systems relative to other aspects of the teacher-selection process now in use, including those dealing with the so-called "imponderables." It will not, however, try to furnish tests of personality, mental adjustment, or other characteristics which must be considered by the local school administrator. Third, the committee will furnish data and lend all possible assistance to promoting research in teacher education and guidance. Recognizing clearly both the importance and complexity of the problems involved in these services, the committee will closely coördinate its activities with the work of the Council's Commission on Teacher Education.

The first examination will be held on March 29-30, 1940, in a dozen or fifteen centers.

Local Units in Ten States.—The Office of Education supplies interesting information concerning local school unit organization in Arizona, Arkansas, California, Illinois, Kentucky, North Carolina, Ohio, Oklahoma, Pennsylvania, and Tennessee. This information on the organization and administration of schools was collected with the coöperation of chief state school officers and state project staffs. Other state agencies such as universities, planning boards, highway departments, tax commissions, the WPA, and PWA lent their assistance. The states surveyed enroll approximately 10,000,000 pupils in about 50,000 schools. Their educational problems are typical of those which may be found in other states.

The survey report, "Local School Unit Organization in 10 States," not only describes in detail the present-day organization and administration of school systems in the several states, but also traces the historical background and development of local school districts. The report presents proposals by the state project staffs for improving public-school organization and administration, and suggests ways for making desirable changes. Other problems treated include inequality of ability to support education and inequality of educational opportunity, the multiplicity of small schools and small districts, failure to assign responsibility for planning district reorganization, obstacles to the reorganiaztion of school units, and relationship of size of school to desirable educational conditions.

The director of the "Local School Units Project" was Henry F. Alves, U. S. Office of Education Specialist in State School Administration, who served in a supervisory and consultative capacity to the state project staffs, each of which assumed responsibility for the preparation of the state project report, available from the state department of education. In addition to the survey report, "Local School Unit Organization in 10 States," the Office of Education has issued a second report on "Principles and Procedures in the Organization of Satisfactory Local School Units."

Carnegie Corporation.—Frederick P. Keppel, President of the Carnegie Corporation of New York, in his annual report announces that grants totaling $4,846,126 were made during the fiscal year

1938-39 by the Corporation to organizations and individuals in the United States and in the British Dominions and Colonies. Of this sum, $697,200 was given for library interests, $1,157,325 for the arts, $1,253,650 for research, studies, and publications, $204,921 for adult education, and $1,533,030 for the general program of the Corporation. The largest grant made during the year was $250,000, given to complete the sum necessary to establish joint academic library services for Vanderbilt University, Peabody College, and Scarritt College at Nashville, Tennessee. Commenting on this grant, Mr. Keppel stated that it "may well provide an example for the other communities having a group of educational institutions."

Grants for scientific research totaled $137,500, and those for educational investigation, $233,550. Grants were made to the amount of $432,500 for the purpose of carrying forward the work of the Institute of International Education, with headquarters in New York, and that of the New Zealand Council for Educational Research. The American Association for Adult Education, the National Occupational Conference, the Committee on Scientific Aids to Learning, and the American Law Institute received a total of $295,400 for work in their respective fields, and $177,000 was voted to the Association of American Colleges, the American Council of Learned Societies, and the National Bureau of Economic Research.

Grants of $500,436 were made for the British Dominions and Colonies. In South Africa these funds have been used chiefly for scientific inquiry; in New Zealand and Australia emphasis has been on educational research, libraries, and social services. In Canada grants have been made primarily to continue research programs in colleges and universities and to forward library extension.

Colorado Study.—The Bureau of Educational Research and Service of the University of Colorado is undertaking a survey study in order to determine which specific problems in learning and teaching are of the most vital concern to teachers in the field. Questionnaires have been sent to approximately 1000 elementary and secondary-school teachers in Colorado, soliciting coöperation in reporting specific difficulties in the classroom on which the findings of educational research and study would be particularly helpful. On the basis of frequency of occurrence, the most important of the problems reported will be selected. Later in the year, the Bureau plans to issue a bulletin in which such selected problems will be analyzed and a synthesis presented of pertinent research findings. The investigation will, incidentally, attempt to determine whether there are significant unsolved problems with which educational research should deal. The director of the Bureau is R. A. Davis.

Studies in Art.—The University of Iowa has completed the third and final volume of its *Studies in the Psychology of Art*, directed by Norman C. Meier, and published by the American Psychological Association. These studies represent a ten-year investigation of the nature and variations of artistic capacity and factors affecting aesthetic response and learning in the art education field. Volume I, published in 1933, was concerned chiefly with the early genetic aspects of artistic capacity, while Volumes II and III concern themselves primarily with the upper grades, high school, and adult levels. The concluding study in Volume III presents for the first time a new theory of artistic aptitude based on all the evidence uncovered either experimentally, statistically, or by case studies during the ten-year period. The program was sponsored

by the Carnegie Foundation for the Advancement of Teaching with funds of the Carnegie Corporation of New York. The procedure involved a merging of psychological, art, and educational viewpoints and was participated in by research assistants representing all three interests under the guidance of Professor Meier.

Iowa Projects.—Iowa State College is carrying forward a number of projects in connection with the Rural Education Sub-section of the Iowa Agricultural Experiment Station.

Uniform County Maps for Each Iowa County.—(Barton Morgan, leader; prepared in coöperation with the Iowa State Planning Board.) It is the purpose of this project to make available to research workers, and others interested in education, uniform county school maps for the 99 counties of Iowa. These maps show county, township, and section lines; school district boundaries; locations of school buildings; types of school districts; roads; and principal streams. They are drawn three-fourths inch to the mile and are in one color. The final tracings of these maps should be completed by June 15, 1940, and Van Dyke negatives prepared by July 1, 1940. The maps will be printed to order, either singly or in quantities, at a price which will cover only the cost of printing and handling charges. (Project 461, State.)

Nature and Extent of the Need for Education in Agriculture for Beginning Farmers in Iowa.—(Barton Morgan and J. A. Starrak, leaders.) The main objectives of this project are: (1) to discover the amount and kind of agricultural education possessed by beginning farmers in Iowa, (2) to secure an estimate of the amount and kind of agricultural education that beginning farmers in Iowa need, and (3) to gather such other data as are fundamental to planning a more adequate program of agricultural education in this state. The data sheet used lists several hundred items under such general headings as the following: personal information, general education, agricultural education, occupational experience, outstanding problems and difficulties encountered, success in meeting agricultural problems, social and personal adjustment, etc. The sample census technique is followed. A field man began work in September of 1938 and has to date interviewed over 600 young farmers located in all sections of the state. The field work was completed about April 1, 1939, and a manuscript will be prepared by March 1, 1940. (Project 590, State.)

Part-time Education in Vocational Agriculture.—(T. E. Sexauer, leader.) The purpose of this study is to discover means for evaluating the effectiveness of part-time schools in agriculture, to determine the factors which will make part-time work effective as an agency in community development, and to ascertain their influence in getting young men started in farming. The data are being collected in the office of the State Board for Vocational Education and in the schools maintaining part-time classes in agriculture. These data include:

a. Expenditures for part-time education in vocational agriculture.
b. Active schools, 1928-1938.
c. Enrollment in courses, ages.
d. Councils, program of work, farm practice.
e. Date of starting and closing sessions per week.
f. Growth, 1928-1938.

The project should be completed by July 1, 1940. (Project 597, State.)

Factors Affecting the Supply and Demand of Home Economics Teachers in Iowa.—(Hester Chadderdon, Florence Fallgatter, and Barton Morgan, leaders.)

The main objectives of this project are: (1) to determine the available personnel for home economics positions in Iowa high schools from 1935-1942, and (2) to determine the nature and size of the demand for high school home economics teachers in Iowa from 1935-1940. Data relative to the numbers of teachers having been or to be graduated in the various institutions will be secured from the registrar's offices of each institution preparing teachers in the state. Records in the offices of the State Superintendent of Public Instruction and of the State Board for Vocational Education will be consulted for the data relative to number of teachers educated outside of Iowa, number of positions in the state, number of new teachers placed, tenure, years of experience, and subject combinations of home economics teachers. The placement records of the institutions preparing teachers will yield information on the subject combinations demanded and the number of teachers securing positions outside the state. A sampling of teachers will be interviewed to determine reasons for short tenure, and the extra-class responsibilities of teachers. A group of superintendents of schools will also be interviewed to secure data on reasons for short tenure, community restrictions on teachers' activities, and any community and administrative desires for more classes in home economics. Interviews will be had with a sampling of college students and college counselors to determine, in so far as possible, reasons for choosing teaching as a profession. These data will be analyzed and compared with those from earlier studies to determine trends and factors affecting the situation. The plan is to complete this project by June 30, 1941.

Graduate Work in the Teachers Colleges of Michigan.—The Board of Regents of the University of Michigan in July, 1938, passed a resolution approving a formal request from the state board of education that the University of Michigan coöperate with the four teachers colleges of the state in planning a program of graduate instruction to be offered at these institutions. According to the action taken, a graduate division has been organized at each of the teachers colleges. The administration of these divisions is the joint responsibility of the teachers colleges and the graduate school of the University. The final authority for the control of these graduate divisions lies with the state board of education and the Board of Regents of the University, but the immediate control is vested in the executive board of the graduate school functioning through two agencies: (1) an advisory council appointed by the president of each of the teachers colleges and (2) the graduate adviser to teachers colleges, Clifford Woody, appointed by the Board of Regents.

According to the coöperative plan of operation the same standards and requirements governing graduate work at the University of Michigan prevail in the graduate divisions of the teachers colleges. The students are admitted to graduate study by the dean of the graduate school at the University; they schedule their program of work according to the regulations of the graduate school and the various departments thereof; they receive their degrees from the University, unless perchance as much as three-fourths of the work for the degree has been completed in courses offered in a graduate division of a teachers college, in which case both the University and the teachers college will be designated on the diploma granted.

Instructional Staff.—The members of the instructional staff of a graduate division of a teachers college and the courses offered are approved by the executive board of the graduate school of

the University of Michigan. The staff members of a division are of three types: members of the regular staff of the University of Michigan, members of the staff of other universities or educational agencies employed by the University of Michigan, and members of the regular staff of the teachers college. Usually the first two types of staff members assume full responsibility for the courses offered, but the third type of staff member, *i.e.*, the staff members of the teachers colleges, have associated with them a member of the University staff known as a sponsor, who offers counsel and advice in planning the course, gives one or more lectures and makes helpful suggestions in the conduct of the course. The relationship of the instructor and the sponsor is identical with that which often prevails at the University when a more experienced instructor coöperates with a less experienced one in the conduct of a course.

Program of Instruction.—The programs of work offered in the graduate divisions during the 1938-39 academic year were naturally limited. Graduate courses were offered at but two of the teachers colleges, Michigan State Normal College and Western State Teachers College. At the former institution, which had already made arrangements to coöperate with the School of Education in offering a program of graduate instruction for teachers of special education, four courses in special education were offered during the first semester. At the latter institution, four graduate courses were given during the second semester: two in education, one in English, and one in history. Seven students, enrolled full-time, elected the courses offered at Michigan State Normal College; 115 students, enrolled part-time, elected the courses at Western State Teachers College.

During the summer of 1939 courses were offered in the graduate divisions of all four of the state teachers colleges. The number of courses offered and the number of students enrolled are as follows:

Institution	*Number of Courses*	*Number of Students*
Central State.....	6	56
Michigan State Normal	7	20
Northern State...	6	57
Western State....	11	154

The courses offered at each institution, except at Michigan State Normal College, were distributed approximately evenly among the academic and professional studies. At Michigan State Normal College, only general professional courses and courses in special education were given. At Northern State Teachers College, all courses were taught by instructors who were members of the graduate faculty of the University of Michigan or were provided by the University. At the other institutions part of the instructors were from the staff of the University or were provided by the University and part were selected from the staffs of the teachers colleges.

During the 1939-40 academic year, programs of courses are being offered in the graduate divisions of all of the teachers colleges except Northern State Teachers College, which to date desires to have courses given in the summer session only. At Central State Teachers College three courses are planned for each semester; at Michigan State Normal College, six courses during the first semester and two during the second semester; and at Western State Teachers College, six courses during each semester. The enrollment during the first semester is as follows: at Central State Teachers College, 36 part-time students; at Michigan State Normal College, 6 full-time and part-time stu-

dents; at Western State Teachers College, 133 part-time students.

General Observations.—The facts cited in the previous sections will give some idea of the nature and extent of the graduate work developed in the coöperative arrangement between the University and the teachers colleges. However, attention should be directed to three features of the plan as put into operation.

1. Only a limited number of graduate courses are offered in each institution. The departments selected for offering graduate work are chosen on the basis of demand for work, library facilities, and availability of suitable instructors. Through limiting the number of courses the institution can concentrate its efforts, in expanding library and other facilities conducive to graduate study, on a few departments.

2. The enrollment of only graduate students in these courses has resulted in the development of a fine *esprit de corps* among the graduate students. Since not many classes were available the enrollment in each class was great enough to stimulate real graduate study. While limiting the enrollment in all courses to graduate students necessarily reduced the number of courses that could be offered and thereby limited the range of choice of courses, one visiting the classes could not help but feel that the disadvantage of this handicap was greatly offset by the advantages of sizeable classes in which conditions conducive to real graduate study prevailed.

3. The policy of coöperation of the University and the teachers colleges in providing the instruction outlined above has been mutually advantageous. The plan has brought about many conferences between the staff members of various departments of the University and of the teachers colleges. The instructors from the University have learned of the problems, facilities, and standards of the teachers colleges; and the instructors from the teachers colleges have learned of similar problems, facilities, and standards emphasized at the University.

From the material presented in the foregoing paragraphs it is evident that considerable progress has been made in the development of graduate work in the teachers colleges. It may seem to some that this development has been rather slow. However, there is general agreement that what progress has been made is real. Many unsolved problems still remain, but they are being attacked. The opinion is ventured that the nature and amount of graduate work done in the teachers colleges during the first one and one-half years are of higher quality and of greater extent than that done in many of the leading universities during the first one and one-half years after graduate work there had been first introduced.

This information was provided by Clifford Woody of the University of Michigan.

Small Rural Schools In the South.—With financial support supplied by the Julius Rosenwald Fund, the Division of Surveys and Field Studies of George Peabody College has closed the second year of an intensive study of one- and two-room rural schools in the South. The study is projected to continue for a number of years. The purposes of the project are to determine, if possible, what critical factors seem to operate in determining the quality of the small rural school in its relation to the improvement of life in the area served by the school. The study has been carried forward by J. E. Brewton of the Division of Surveys and Field Studies and Horace M. Bond of Fisk University. Preliminary reports are nearing completion. The extension of the investigation will include intensive work in selected counties and in selected insti-

tutions engaged in training teachers for small rural schools.

Theses on Micro-films.—The School of Education has joined with the other schools of Pennsylvania State College in making its dissertations available on micro-films. These are produced by University Microfilms, Ann Arbor, Michigan. This plan includes:

1. A volume of four or five page abstracts of all Doctors' dissertations of the university, published by the university.
2. A film of each complete dissertation, copies of which may be purchased from the producer, or borrowed from the Library of Pennsylvania State College.
3. A volume published by University Microfilms annually, listing all of the dissertations of the coöperating universities.

Studies of Reading at Syracuse University.—M. E. Troyer is doing research on the development of films for remedial treatment in the fields of reading and arithmetic. The Educational Clinic, of which Virgil E. Herrick has charge, is concentrating its research in the field of reading. Studies under way are concerned with: visual factors in relation to reading, social adjustment of children in grades one to three and reading success, the background of experience and its relation to meaning for children in the elementary school, and emotional maladjustment of children in relation to progress in reading.

Coöperative Test Service.—The primary function of any general testing program is to provide accurate and meaningful measures of the attainments of students in basic subjects and in broad fields. These measures furnish information that is essential for intelligent student self-guidance, for educational and vocational counseling, for adapting instruction to individual differences in backgrounds and needs, for improving supervisory procedures, and for evaluating curricula, teaching methods, and other general features of the educational program.

The preparation of a coördinated battery of tests for high school use is a large undertaking. The Coöperative Test Service, a nonprofit agency of the American Council on Education, was established primarily for the purpose of developing such tests. Examinations in several general fields and in nearly all the major academic subjects of high school instruction are now available. Two years ago the first editions of the Revised Series Coöperative Achievement Tests were issued. Most of these are 40-minute examinations. Because of the care used in selecting for inclusion only the best items from long lists, these tests have adequate reliability for individual diagnosis in most cases, and at the same time they are short enough to be administered in a single class period of the usual length. Scaled Scores and differentiated norms make possible a variety of meaningful comparisons between the results of one test and another as well as between those of one form or edition and another, and facilitate the use of profile charts and permanent cumulative record forms. These examinations and the accompanying interpretative materials are now offered as tested instruments for annual measurement programs at the secondary level.

JOU

Volume X

Ed
tention
supplies

THIS a
reading pu
the series t
February,
reader wit
with the m
should be
of special

Major
in the scie
The numb
any previo
reported c

An anal
is still dire
school leve
issues most
of reading
of words t
phonics, m
reading, th
cepts inclu
reading, m
influence o

Overlap